

LA MODERACIÓN DE CONTENIDOS DESDE UNA PERSPECTIVA INTERAMERICANA





AlSur









LA MODERACIÓN DE CONTENIDOS DESDE UNA PERSPECTIVA INTERAMERICANA

Por: Vladimir Alexei Chorny Elizalde, Luis Fernando García Muñoz y Grecia Elizabeth Macias Llanas

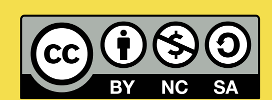
Contribución de Al Sur al Diálogo de las Américas sobre Libertad de Expresión en internet para recibir insumos para la elaboración de estándares en la materia, lanzado por la Relatoría Especial para la Libertad de Expresión (RELE) de la Comisión Interamericana de Derechos Humanos (CIDH).

"Al Sur" es un consorcio de organizaciones que trabajan en la sociedad civil y en el ámbito académico en América Latina y que buscan con su trabajo conjunto fortalecer los derechos humanos en el entorno digital de la región.

Ciudad de México. México, Marzo 2022



Diseño Editorial: Gibrán Aquino Pineda





ÍNDICE

A. EL DERECHO A LA LIBERTAD DE EXPRESIÓN EN EL SISTEMA INTERAMERICANO

a. Principios generales

El derecho a la libertad de expresión, reconocido en el artículo 13 de la Convención Americana sobre Derechos Humanos (CADH) posee un tratamiento especial dentro del Sistema Interamericano de Derechos Humanos (en adelante SIDH). Es por ello que la Corte Interamericana de Derechos Humanos (en adelante "Corte IDH") ha resaltado la preponderancia de la libertad de expresión y reiterado que ésta constituye la piedra angular de las sociedades democráticas y que es también una condición esencial para que estén suficientemente informadas.[1]

Derivado de lo anterior, la Corte IDH ha destacado que *"[...] las garantías de la libertad de expresión contenidas en la [CADH] fueron diseñadas para ser las más generosas y para reducir al mínimum las restricciones a la libre circulación de las ideas"*.[2] Esto trae como consecuencia que las restricciones de otros sistemas -como el europeo- no puedan ser aplicadas directamente en el marco interamericano.

La Comisión Interamericana de Derechos Humanos (CIDH) y la Corte IDH han sido clave para dotar de contenido a la libertad de expresión. Uno de los aspectos particularmente trascendentales emanados de la doctrina interamericana es el reconocimiento de la doble dimensión, individual y colectiva, que posee el derecho a la libertad de expresión.[3]

[1] Corte IDH. Serie C No. 73. Caso "La Última Tentación de Cristo" (Olmedo Bustos y otros) vs. Chile. Fondo y Reparaciones y Costas. Sentencia de 5 de febrero de 2001, párr. 68.

[2] Corte IDH. Serie A No. 5. La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 de la Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, párr. 52.

[3] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre el derecho a la libertad de expresión. OEA/Ser.L/V/II CIDH/RELE/INF.2/09, 30 diciembre 2009, párr. 2

La doble dimensión "requiere, por un lado, que nadie sea arbitrariamente menoscabado o impedido de manifestar su propio pensamiento y representa, por tanto, un derecho de cada individuo; pero implica también, por otro lado, un derecho colectivo a recibir cualquier información y a conocer la expresión del pensamiento ajeno".[4] Esto quiere decir que las personas tenemos el derecho de exponer nuestro punto de vista y de recibir y conocer el de otras personas,[5] por lo que si un acto de un Estado afecta o restringe la dimensión individual del derecho en cabeza del emisor, afecta de igual forma y en la misma medida la dimensión social en cabeza del receptor.[6]

4 Corte IDH. Serie C No. 107. Caso Herrera Ulloa vs. Costa Rica. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 2 de julio de 2004, párr 108; Serie C No 111. Caso Ricardo Canese vs. Paraguay. Fondo, Reparaciones y Costas. Sentencia de 31 de agosto de 2004, párr.77.; y I Serie C No. 74. Caso Ivcher Bronstein vs. Perú. Reparaciones y Costas. Sentencia de 6 de febrero de 2001, párr. 146.

5 Corte IDH, Opinión Consultiva OC-5/85. La colegiación obligatoria de periodistas (arts. 13 y 29 Convención Americana sobre Derechos Humanos). 13 de noviembre de 1985, serie A núm. 5, párr. 30; CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre el derecho a la libertad de expresión. OEA/Ser.L/V/II CIDH/RELE/INF.2/09, 30 diciembre 2009, párr. 13; Corte IDH, caso Kimel vs. Argentina. Fondo, Reparaciones y Costas. Sentencia de 2 de mayo de 2008, serie C núm. 177, párr. 53; Corte IDH, caso Claude Reyes y otros vs. Chile. Fondo, Reparaciones y Costas. Sentencia de 19 de septiembre de 2006, serie C núm. 151, párr. 75; Corte IDH, caso López Álvarez vs. Honduras. Fondo, Reparaciones y Costas. Sentencia de 1 de febrero de 2006, serie C núm. 141, párr. 163; Corte IDH, caso Herrera Ulloa vs. Costa Rica. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 2 de julio de 2004, serie C núm. 107, párr. 108; Corte IDH, caso Ivcher Bronstein vs. Perú. Fondo, Reparaciones y Costas. Sentencia de 6 de febrero de 2001, serie C núm. 74, párr. 146; Corte IDH, caso Ricardo Canese vs. Paraguay. Fondo, Reparaciones y Costas. Sentencia de 31 de agosto de 2004, serie C núm. 111, párr. 77; Corte IDH, caso «La Última Tentación de Cristo» (Olmedo Bustos y otros) vs. Chile. Fondo, Reparaciones y Costas. Sentencia de 5 de febrero de 2001, serie C núm. 73, párr. 64; CIDH, Informe núm. 130/99, caso núm. 11.740, Víctor Manuel Oropeza, México, 19 de noviembre de 1999, párr. 51; CIDH, Informe núm. 11/96, caso núm. 11.230, Francisco Martorell, Chile, 3 de mayo de 1996, párr. 53.

6 Corte IDH, Opinión Consultiva OC-5/85. La colegiación obligatoria de periodistas (arts. 13 y 29 Convención Americana sobre Derechos Humanos). 13 de noviembre de 1985, serie A núm. 5, párr. 33; CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre el derecho a la libertad de expresión. OEA/Ser.L/V/II CIDH/RELE/INF.2/09, 30 diciembre 2009, párr. 15; Corte IDH, caso Palamara Iribarne vs. Chile. Fondo, Reparaciones y Costas. Sentencia de 22 de noviembre de 2005, serie C núm. 135, párr. 107; Corte IDH, caso Ricardo Canese vs. Paraguay. Fondo, Reparaciones y Costas. Sentencia de 31 de agosto de 2004, serie C núm. 111, párr. 81; CIDH, alegatos ante la Corte IDH en el caso Herrera Ulloa vs. Costa Rica, transcritos en: Corte IDH, caso Herrera Ulloa vs. Costa Rica. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 2 de julio de 2004, serie C núm. 107, párr. 101-1-a; CIDH, Informe de fondo núm. 90/05, caso núm. 12.142, Alejandra Marcela Matus Acuña, Chile, 24 de octubre de 2005, párr. 39.

Cuando la Corte IDH, por ejemplo, ha discutido sobre controlar las noticias falsas, ha sostenido esta relación entre las dos dimensiones de la libertad de expresión:

> *"...no sería lícito invocar el derecho de la sociedad a estar verazmente informada para fundamentar un régimen de censura previa supuestamente destinado a eliminar las informaciones que serían falsas a criterio del censor. Como tampoco sería admisible que, sobre la base del derecho a difundir informaciones e ideas, se constituyan monopolios públicos o privados sobre los medios de comunicación para intentar moldear la opinión pública según un solo punto de vista".*[7]

El sistema interamericano de derechos humanos establece desde aquí un régimen de obligaciones y responsabilidades que tiene distintos alcances y que exige distintas acciones a los distintos sujetos que pueden involucrarse con los derechos contenidos en la CADH. Para garantizar el cumplimiento de la libertad de expresión, el marco interamericano obliga al Estado a tomar medidas negativas o de abstención frente a los derechos, por ejemplo no legislando en contra de la libertad de expresión, pero también a tomar medidas positivas para hacer que el derecho sea verdaderamente efectivo, por ejemplo al llevar a cabo acciones frente a actores particulares o privados para impedir que sus acciones lesionen alguna dimensión del derecho[8] (como sucede en los casos de medidas antimonopólicas o anti-concentración de los medios de comunicación).

[7] Corte IDH, Opinión Consultiva OC-5/85. La colegiación obligatoria de periodistas (arts. 13 y 29 Convención Americana sobre Derechos Humanos). 13 de noviembre de 1985, serie A núm. 5, párr. 33; CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre el derecho a la libertad de expresión. OEA/Ser.L/V/II CIDH/RELE/INF.2/09, 30 diciembre 2009, párr. 1. Las noticias falsas no son, en el fondo, más que desinformación o promoción de información inexacta.

[8] La relación de las obligaciones estatales frente a las acciones de los sujetos no estatales ha sido reconocida en distintas ocasiones por la Corte IDH. Dos ejemplos son el caso *Juan Humberto Sánchez* y el caso *Maritza Urrutia*, donde la Corte señaló explícitamente que el marco de la CADH reconoce deberes positivos "que imponen exigencias específicas tanto a los agentes del Estado como a terceros que actúen con su tolerancia o anuencia y que sean responsables de la detención" (párrafos 81 y 71, correspondientemente). Al respecto ver: Corte IDH. Caso Juan Humberto Sánchez Vs. Honduras, Sentencia de 7 de junio de 2003, Excepción Preliminar, Fondo, Reparaciones y Costas; Corte IDH. Caso Maritza Urrutia Vs. Guatemala, Sentencia de 27 de noviembre de 2003, Fondo, Reparaciones y Costas.



La Corte IDH reconoce una tipología de obligaciones estatales que se dividen primero, en general, en las obligaciones de respetar y las obligaciones de garantizar los derechos, dentro de las que después encontramos, en particular, las obligaciones de proteger, de crear instituciones para investigar, sancionar y reparar, y de promover los derechos humanos (estas últimas tres incluidas conceptualmente dentro de la obligación de garantizar los derechos).

En cuanto a las obligaciones negativas, el ejemplo más claro está en la obligación de respetar, que implica que las autoridades no lleven a cabo acciones que lesionen derechos humanos: esta dimensión reconoce la visión clásica de los derechos como esferas individuales que mantienen alejado al Estado y restringen su poder frente a los individuos.[9]

En cuanto a las obligaciones positivas, tanto la obligación de proteger (consistente en asegurarse de que las personas no vean violados sus derechos ni por las autoridades ni por sujetos particulares) como la de garantizar (consistente en la adopción de medidas estatales -todas las necesarias- para para crear las condiciones que permitan gozar efectivamente de los derechos), desdoblan una dimensión de acciones que salen de la visión clásica del Estado y que le dan un rol activo ya no solo como actor central para el ejercicio pleno de los derechos, sino también para encargarse que otros sujetos no estatales no los obstaculicen y cumplan con ellos.[10]

El marco de obligaciones positivas pone en el foco el rol que los sujetos privados tienen frente al respeto y la garantía de los derechos humanos. En el tema que ocupa este trabajo, está en el centro el trabajo de las empresas y plataformas que tienen una posibilidad -o poder- real para alterar el flujo informativo y afectar derechos como el acceso a la información y la libertad de expresión.[11] Un ejemplo claro de este tipo de obligaciones puede encon-

[9] Corte IDH. Caso Velásquez Rodríguez vs. Honduras, Fondo, sentencia de 29 de julio de 1988, serie C núm. 4, párr. 165.

[10] Salazar Ugarte, Pedro. *La Reforma Constitucional de Derechos Humanos. Una Guía Conceptual*, México, Instituto Belisario Domínguez, 2014, pp. 112-117. El carácter positivo de estas obligaciones estatales exige una acción efectiva en contra de particulares, que va desde la toma de medidas -por ejemplo- frente a una empresa que contamina el medio ambiente, hasta aquellas necesarias para que las empresas respeten la privacidad y la libertad de expresión (cuestiones particularmente relevantes para el trabajo que nos ocupa).

trarse en el marco de los *Principios Rectores sobre las Empresas y Derechos Humanos*, que es probablemente el instrumento más importante en la materia y que señala de manera puntual los deberes de las empresas y otros sujetos privados sobre el respeto, la protección y la reparación por violaciones a derechos humanos. Dicho marco no sólo reconoce la obligación general de respetar los derechos, sino también las obligaciones específicas de *actuar con diligencia debida* y de *ser transparentes*, así como la obligación de *reparar* por violaciones a los derechos dentro del marco de sus competencias.[12]

Es fácil pensar en casos donde las empresas tengan estas obligaciones frente a la libertad de expresión en Internet. La moderación de contenidos es probablemente uno de los más interesantes para pensar las obligaciones y el SIDH es particularmente relevante para pensar estas relaciones porque reconce que la libertad de expresión debe garantizarse a cualquier persona sin discriminación alguna, dentro de un marco complejo de deberes y responsabilidades que dependen de la situación concreta en la que se ejerce el derecho y del procedimiento técnico utilizado para manifestar y difundir esa expresión.[13]

La libertad de expresión tiene entonces una flexibilidad particular que debe ser tomada en cuenta. Un buen ejemplo de su flexibilidad puede encontrarse en lo que se conoce como los discursos especialmente protegidos y los discursos que no cuentan con la defensa reforzada del derecho a la libertad de expresión.

[11] La discusión sobre la "eficacia horizontal" de los derechos trae a la luz un tema que suele desplazarse o minimizarse en la crítica académica y política: el hecho de que algunas empreas tienen las capacidades y el poder suficiente para ser consideradas como sujetos obligados frente al conjunto de derechos humanos con el que se relacionan; es decir, que algunos sujetos no estatales no sólo pueden -y en efecto lo hacen- violar derechos humanos, sino que además están obligados de manera concreta a llevar cierto tipo de acciones para respetarlos y garantizarlos. Al respecto ver: Ziemele, Ineta. "Human Rights Violations by Private Persons and Entities: The Case-Law of International Human Rights Courts and Monitoring Bodies", European University Institute-Academy of European Law, EUI AEL; 2009/08; Nolan, Aoife (2014), "Holding non-state actors to account for constitutional economic and social rights violations: Experiences and lessons from South Africa and Ireland", I•CON (2014), Vol. 12 No. 1, pp. 61-93; Chorny, Vladimir. "La violación de derechos humanos por sujetos no estatales: una visión completa de los derechos". *Revista Latinoamericana de Filosofía Política*, Marzo de 2018.

[12] Consejo de Derechos Humanos. *Principios Rectores sobre las Empresas y Derechos Humanos*, Organización de las Naciones Unidas, 2011.

[13] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre el derecho a la libertad de expresión. OEA/Ser.L/V/II CIDH/RELE/INF.2/09, 30 diciembre 2009, párr. 18.



b. Discursos protegidos y discursos especialmente protegidos

Las características propias del derecho a la libertad de expresión protegen los distintos tipos de expresiones independientemente de su forma, contenido o medio de comunicación, tal como lo contempla el artículo 13 de la CADH, que señala que:

> *"Toda persona tiene derecho a la libertad de pensamiento y de expresión. Este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de toda índole, sin consideración de fronteras, ya sea oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento de su elección".*[14]

Todas las expresiones (orales, escritas, artísticas, etc.) tienen una protección "en principio" (a la que suele referirse como cobertura ab initio), lo que significa que existe una presunción de que todas las expresiones se encuentran protegidas incluso si estas pueden llegar a ser consideradas chocantes, ofensivas o perturbadoras. Como regla general, se trata de un derecho sujeto a un régimen muy limitado de excepciones, expresa y puntualmente definidas en el derecho internacional mediante prohibiciones concretas y específicas.[15]

La obligación que los Estados tienen de ser neutrales ante los contenidos que son expresados en el marco de este derecho es resultado de la presunción de cobertura y es también un efecto de la necesidad de garantizar que, en principio, no existan personas, grupos, ideas o medios de expresión excluidos a priori del debate público.[16] Esta máxima de la libertad de expresión es la que trae como resultado la prohibición de la censura previa.

[14] Convención Americana sobre Derechos Humanos, San José, Costa Rica, 12 al 22 de noviembre de 1969, Organización de los Estados Americanos.

[15] Centro de Estudios de Derecho, Justicia y Sociedad, Dejusticia. *El derecho a la libertad de expresión*, Colombia, 2017, p. 59. Lo cual no significa que esa presunción siempre se sostenga ni que se trate de un derecho absoluto, sino que las limitaciones del derecho, también como regla general, deben ser posteriores a que la manifestación se haya expresado.

[16] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 30.

La doctrina interamericana ha clasificado las expresiones especialmente protegidas, a grandes rasgos, en tres tipos de discurso:

1. Discurso político y sobre asuntos de interés público

En una sociedad democrática, la importancia de la discusión pública relacionada con el ámbito político y los asuntos de interés general acortan el margen de restricciones legítimas a la crítica política y a las manifestaciones relacionadas con cuestiones de interés público. Tanto la CIDH como la Corte IDH han impulsado esta doctrina al explicar que el funcionamiento de la democracia exige el mayor nivel posible de discusión pública sobre el funcionamiento de la sociedad y del Estado en todos sus aspectos, esto es, sobre los asuntos de interés público. De allí que el adecuado desenvolvimiento de la democracia requiera la mayor circulación de informes, opiniones e ideas sobre asuntos de esta índole.[17]

La Convención Americana sobre Derechos Humanos reconoce la protección ampliada para este tipo de expresiones, cosa que ha sido sistemáticamente reiterada por su órgano interpretativo principal (la Corte IDH). La protección ampliada implica que existen obligaciones claras para que los Estados se abstengan rigurosamente de establecer límites a las formas de manifestación de expresiones, por un lado, pero también para explicar que las personas que participen de la discusión pública deben tener un mayor umbral de tolerancia a la crítica.[18]

2. Discurso sobre funcionarios públicos y sobre candidatos a ocupar cargos públicos y otras figuras públicas

Cuando las expresiones de las personas se dirijan a funcionarios públicos, a personas particulares involucradas voluntariamente en asuntos públicos

[17] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párrs. 57 y 87; Corte I.D.H., Caso Claude Reyes y otros Vs. Chile. Sentencia de 19 de septiembre de 2006. Serie C No. 151, párrs. 84, 86 y 87; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 83; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 127.

[18] Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 83; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 125; CIDH. Alegatos ante la Corte Interamericana en el caso Herrera Ulloa Vs. Costa Rica. Transcritos en: Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 101.2.c)



o a candidatos a ocupar cargos públicos,[19] se repite la fórmula que lleva a reconocer un umbral mayor de tolerancia ante la crítica,[20] por lo que las obligaciones de abstención por parte del Estado (en cuanto a las restricciones y limitaciones al ejercicio del derecho) también están presentes.

Todos estos grupos de sujetos (personas funcionarias, candidatas y particulares que son parte de asuntos públicos) participan de manera voluntaria bajo este régimen de escrutinio público fuerte, en el que la crítica a sus acciones por parte del público funciona como un mecanismo de rendición de cuentas que es parte de la noción más amplia de control democrático. Los controles democráticos están justificados para mantener bajo revisión al ejercicio del poder público, a través de la obligación de transparencia y de máxima publicidad que ordena todas las acciones del Estado.[21]

Si las personas participan en el ámbito de lo público, no sólo es esperable que estén sujetas a un escrutinio público fuerte, sino que también es comprensible un mayor nivel de crítica porque la capacidad de respuesta de estos sujetos también es mayor, sea por su convocatoria pública, su influencia social o su acceso a los medios de comunicación (tal como reconoció también la Corte IDH en el Caso Tristán Donoso vs. Panamá, por mencionar tan solo un ejemplo).[22]

[19] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párr. 86; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 82.

[20] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párrs. 86-88; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párrs. 83-84; Corte I.D.H., Caso "La Última Tentación de Cristo" (Olmedo Bustos y otros) Vs. Chile. Sentencia de 5 de febrero de 2001. Serie C No. 73, párr. 69; Corte I.D.H., Caso Ivcher Bronstein Vs. Perú. Sentencia de 6 de febrero de 2001. Serie C No. 74, párrs. 152 y 155; Corte I.D.H., Caso Ricardo Canese. Sentencia de 31 de agosto de 2004. Serie C No. 111, párr. 83; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107., párrs. 125 a 129; Corte I.D.H., Caso Claude Reyes y otros. Sentencia de 19 de septiembre de 2006, Serie C No. 151, párr. 8.

[21] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 40.

[22] Corte I.D.H., Caso Tristán Donoso Vs. Panamá. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 27 de enero de 2009 Serie C No. 193, párr. 122.

3. Discursos que expresan elementos esenciales de la identidad o dignidad personales

La libertad de expresión no solo es un derecho en sí mismo sino que también es un derecho habilitador de otros derechos y que funciona como herramienta para fortalecer la identidad y dignidad de las personas. Su carácter de habilitador y de condición necesaria para aquellos derechos como la identidad personal hace que los discursos relacionados con este tipo de expresiones también estén protegidos de manera reforzada.

Un ejemplo que ha sido reiterado en este sentido dentro del SIDH es el de los derechos de los pueblos indígenas a expresarse y a recibir información en el idioma que forma su identidad, ya que la lengua propia es uno de los elementos centrales a tomar en cuenta en la conformación de la identidad de las personas y los grupos, y es a partir de ella que estos pueden expresarse. Pero además, en el caso de los pueblos indígenas, la transmisión de su cultura y de la cosmovisión que los diferencia del resto de la población no indígena también pasa por el uso del lenguaje y de la conformación de su identidad cultural (a partir de él).[23]

Otro ejemplo es el de la reivindicación de la libertad de pensamiento y de expresión en cuestiones relativas a la diversidad sexual, al reconocimiento de la identidad sexual o de género y a la importancia que estos derechos tienen para evitar cualquier tipo de censura sobre las expresiones relacionadas con estas cuestiones. En su Opinión Consultiva 24/17, la Corte señaló como casos de censura indirecta cuando un sistema jurídico no reconocía la identidad de género y cuando eran castigadas o censuradas (aún indirectamente) las expresiones de género que se apartaban de los estándares cisnormativos o heteronormativos, ya que en esos casos se transmitía el mensaje de que quienes quedaban por fuera de los estándares "tradicionales" no recibían la misma consideración y respeto ni la misma protección legal y el reconocimiento de sus derechos. [24]

[23] Corte I.D.H., Caso López Álvarez Vs. Honduras. Sentencia del 1º de febrero de 2006. Serie C No. 141. párr. 169.

[24] Corte I.D.H., Identidad de género, e igualdad y no discriminación a parejas del mismo sexo. Opinión Consultiva OC-24/17 del 25 de noviembre de 2017, Serie A No. 24; Comisión Interamericana de Derechos Humanos, Observaciones presentadas por la Comisión el 14 de febrero de 2017, párr. 49. Véase, en el mismo sentido, Naciones Unidas, Comité de los Derechos del Niño, Observación general núm. 20 (2016) sobre la efectividad de los derechos del niño durante la adolescencia,



c. Discursos no protegidos

En el otro extremo del ejercicio de la libertad de expresión, está otra serie de expresiones que están abierta y tajantemente prohibidas y que no cuentan con la cobertura que se da a las expresiones protegidas. En estos casos es posible tomar medidas más restrictivas e incluso censurar previamente situaciones excepcionales reservadas a casos muy concretos, como el que representa el abuso sexual de menores, también conocido como "pornografia infantil".

De conformidad con el derecho internacional de los derechos humanos y concretamente lo previsto en el artículo 13 de la CADH. Existen tres tipos de discursos que no se encuentran protegidos por la libertad de expresión.

1. Incitación a la violencia

El ejercicio de la libertad de expresión puede ser sancionado cuando se lleva a cabo de manera abusiva. Tanto a nivel doctrinal como a nivel jurisprudencial existe un amplio acuerdo de que cuando se realiza la conducta de incitar a la violencia (incitar a cometer crímenes, a romper el orden público o afectar la seguridad nacional), es aceptable establecer sanciones incluso desde el ámbito del derecho penal.

El artículo 13.5 de la CADH señala concretamente que: "Estará prohibida por la ley toda propaganda en favor de la guerra y toda apología del odio nacional, racial o religioso que constituyan incitaciones a la violencia o cualquier otra acción ilegal similar contra cualquier persona o grupo de personas, por ningún motivo, inclusive los de raza, color, religión, idioma u origen nacional".

Sin embargo, en estos casos, debe ser claro que existe una prueba de carácter actual, cierta, objetiva y contundente de que la alegada conducta de incitación no era la simple manifestación de una opinión (por dura, injusta y perturbadora que fuera) y que tenía no solo una intención clara

6 de diciembre de 2016, CRC/C/GC/20, para. 34, y Oficina del Alto Comisionado de las Naciones Unidas, Living Free & Equals. What States are doing to tackle violence and discrimination against lesbian, gay, bisexual, transgender and intersex people, Nueva York y Ginebra, 2016, HR/PUB/16/3, págs. 86 y 87.

de cometer un crimen sino también la posibilidad actual, real y efectiva de lograr sus objetivos.[25]

2. La incitación directa y pública al genocidio

De la misma manera que con la incitación a la violencia, la incitación directa y pública al genocidio está prohibida tanto en los estándares internacionales e interamericanos sobre libertad de expresión, como en otros instrumentos internacionales especializados que regulan las acciones relacionadas al crimen de genocidio, tal como la Convención para la Prevención y la Sanción del Delito de Genocidio.[26]

En el texto de dicha convención se considera como genocidio a la matanza de miembros de un grupo, la lesión grave de su integridad física o mental, el sometimiento intencional del grupo a condiciones de existencia que acarreen su destrucción física (total o parcial), las medidas destinadas a evitar nacimientos en el grupo y el traslado forzoso de niñas y niños de dicho grupo, cuando dichas conductas se realicen con la intención de destruir, total o parcialmente a un grupo nacional, étnico, racial o religioso.[27] La conducta que es castigable penalmente, en este sentido, es la "instigación directa y pública a cometer[lo]".[28]

3. El abuso sexual de menores (pornografía infantil)

Existe un consenso internacional para prohibir de forma categórica la "pornografía infantil". El interés superior de las niñas, los niños y adolescentes es lesionado inevitablemente por un discurso que es violento hacia ellos y que viola sus derechos,[29] los cuáles son reconocidos y protegidos por el Estado (así como por la sociedad y la familia).[30]

[25] Corte IDH. La Colegiación Obligatoria de Periodistas (arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985. Serie A No. 5, párr. 77.; CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 58.

[26] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 59.

[27] Convención para la Prevención y la Sanción del Delito del Genocidio, AGNU, 09 de diciembre de 1948, artículo II.

[28] Convención para la Prevención y la Sanción del Delito del Genocidio, AGNU, 09 de diciembre de 1948, artículo III, inciso c).

[29] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 59.

[30] CADH. Artículo 19.



La “pornografía infantil” es rechazada a nivel internacional por considerarse una forma de explotación y abuso sexuales, y se castiga tanto la incitación y la coacción en cualquier acto de índole sexual como en cualquier espectáculo o material pornográfico.[31]

El Sistema Interamericano de Derechos Humanos es muy claro: salvo estas limitaciones, concretas y reservadas para los casos extremadamente graves del ejercicio ilegítimo del derecho a expresarse, todas las expresiones deben someterse al régimen de responsabilidades ulteriores que privilegia y protege la difusión de las mismas y que establece un sistema de sanciones posteriores que pueden fijarse en otros casos en los que el ejercicio del derecho afecte intereses legítimos de otras personas, pero siempre por fuera y más allá del régimen de censura.

d. Limitaciones a la libertad de expresión

Hasta este punto se ha dejado claro que todas las expresiones tienen una presunción de protección bajo la libertad de expresión, a reserva de las tres excepciones analizadas con anterioridad. La protección reforzada de la libertad de expresión se justifica porque las sociedades democráticas defienden la convicción de que es valioso tener la mayor cantidad de elementos para pensar, informarse y expresar las ideas y sentimientos de las personas en el ámbito público. Sin embargo, ningún derecho es absoluto y pueden existir situaciones bajo las cuales puede limitarse la libertad de expresión; limitaciones que deben seguir una serie de condiciones estrictas para considerarlas conforme a derecho.

Lo anterior significa que para que una limitación a la libertad de expresión se considere “legítima” o “válida” (según hagamos un juicio sobre su dimensión política o su dimensión jurídica), deberá estar sujeta a las condiciones establecidas por el derecho interamericano (del SIDH). La regla sobre los límites a la libertad de expresión funciona tanto frente a las autoridades estatales (de todas las ramas del poder público) como hacia los sujetos no estatales (sea que se trate de particulares realizando funciones estatales o con financiamiento público, o que lo hagan por sí mismos).[32]

[31] Convención sobre los Derechos del Niño, artículo 34.
[32] Centro de Estudios de Derecho, Justicia y Sociedad, Dejusticia. *El derecho a la libertad de expresión*, Colombia, 2017, p. 96.

Como señalamos previamente, el sistema de protecciones y límites de este derecho tiene como base la prohibición de la censura previa y como regla general el régimen de responsabilidades ulteriores opera para todas las expresiones (salvo las excepciones señaladas). La doble faceta de este derecho debe su razón de ser a que, en una sociedad democrática, es aceptable considerar discursos que son valiosos para su sostenimiento y mejoramiento, y que hay otros que se consideren disvaliosos por significar justo lo contrario.[33]

Para la regla de la prohibición de censura previa, no es aceptable establecer ningún requisito, condición o autorización previa a las expresiones (salvo en los casos de discursos no protegidos, dado que se trata de intereses especialmente vulnerables que justifican un resguardo calificado); para la regla de las responsabilidades ulteriores, existe un parámetro conocido como el *test tripartito*, a través del cual podemos analizar si una limitación particular a la libertad de expresión es *válida* o no.

A partir del artículo 13.2 de la CADH, el sistema interamericano ha desarrollado -de la mano de la Corte IDH- un estándar y un método que sirve para saber qué pasos deben de seguirse si se busca limitar de manera legítima el derecho a la libertad de expresión (y, en el aspecto jurídico, para determinar si dicha limitación es válida o no, si viola el derecho o no, etc.).[34]

En primer lugar (paso 1), la limitación debe estar plasmada previamente en una ley (formal y materialmente), y debe estar ahí definida de manera expresa y taxativa en relación a alguno de los fines legítimos reconocidos por la propia CADH (los llamados "objetivos democráticos").[35] Los objetivos democráticos están delimitados por la propia CADH y corresponden a los derechos o la reputación de los demás, la protección de la seguridad nacional, el orden público, la salud o la moral públicas. La interpretación de estos objetivos debe ser siempre "democrática" en el sentido de que estos intereses no deben to-

[33] Salazar Ugarte, Pedro y Gutiérrez Rivas, Rodrigo. *El derecho a la libertad de expresión frente al derecho a la no discriminación*, México, Instituto de investigaciones Jurídicas-UNAM, 2008, p. 28.
[34] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 67.
[35] CIDH. Alegatos ante la Corte Interamericana en el caso Ricardo Canese Vs. Paraguay. Transcritos en: Corte I.D.H., Caso Ricardo Canese Vs. Paraguay. Sentencia de 31 de agosto de 2004. Serie C No. 111, párrs. 72. s) a 72.u).



marse como superiores a la libertad de expresión sino como articulables con ella para maximizarla y fortalecer el sistema democrático como un todo.[36]

Cuando se trata de los derechos de otras personas, la interpretación adecuada debe partir primero de que sea claro que dichos derechos se hayan lesionado o amenazado, para después evaluar los grados en que esto es así frente al peso privilegiado de la libertad de expresión.[37]

La Corte IDH ha sido muy clara al señalar que es contradictorio "invocar una restricción a la libertad de expresión como un medio para garantizarla, porque es desconocer el carácter radical y primario de ese derecho como inherente a cada ser humano individualmente considerado, aunque atributo, igualmente, de la sociedad en su conjunto".[38] También que no es aceptable exigir que el ejercicio de la libertad de expresión esté atado a una condición de veracidad, porque si esto fuera así se abriría una puerta de abusos sobre los controles de la información que afectaría el derecho de acceso a la información de todas las personas.[39]

Cuando se trata del interés por el "orden público", la Corte ha tratado este principio como "las condiciones que aseguran el funcionamiento armónico y normal de las instituciones sobre la base de un sistema coherente de valores y principios",[40] pero que a su vez, este principio requiere que

[36] Opinión Consultiva OC-5/85 de la Corte IDH, La Colegiación Obligatoria de Periodistas, de 13 de noviembre de 1985. Además de esto, la CIDH ha desarrollado el criterio para señalar que las limitaciones deben ser compatibles con el "principio democrático", que implica que deben: i) incorporar las exigencias justas de una sociedad democrática; ii) ser compatibles con la preservación y el desarrollo de las sociedades democráticas de acuerdo a los artículos 29 y 32 de la CADH; y iii) ser interpretadas con referencia a las necesidades legítimas de las sociedades y las instituciones democráticas. Al respecto ver: CIDH. Informe Anual 2009. Informe de la Relatoría Especial para la Libertad de Expresión. Capítulo III (Marco Jurídico Interamericano del Derecho a la Libertad de Expresión). OEA/Ser.L/V/II. Doc. 51. 30 de diciembre de 2009. Párr. 67.

[37] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 77.

[38] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 7

[39] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 77.

[40] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 64.

"se garanticen las mayores posibilidades de circulación de noticias, ideas y opiniones, así como el más amplio acceso a la información por parte de la sociedad en su conjunto".[41]

Cuando se trata de la "seguridad nacional" la lógica ha sido la misma. Una interpretación abarcativa de dicha limitación es incompatible con las sociedades democráticas. Por el contrario, las democracias modernas requieren que este interés se interprete a la luz del carácter primario de la libertad de expresión y de la necesidad de tener la mayor y mejor información sobre los asuntos públicos de interés de la sociedad disponible.

En el caso de Internet, este punto ha sido enfatizado particularmente a la luz de la información relacionada con los programas de vigilancia estatal (y sobre las reservas relacionadas a ésta), en donde la Relatoría para la Libertad de Expresión de la CIDH ha sido clara al decir que no es legítimo limitar esta información bajo la categoría de seguridad nacional cuando se intercepte, capture o utilice información privada de disidentes, periodistas o defensores de derechos humanos con finalidades políticas o para evitar o comprometer sus investigaciones o denuncias.[42]

La limitación establecida en la ley debe además ser clara y precisa. Todas las restricciones que no cumplan con estos requisitos implican la vulneración del primer elemento del test tripartito y se consideran contrarias al marco interamericano, particularmente porque abren un margen de discrecionalidad demasiado amplio para las autoridades y porque habilitan un espacio de arbitrariedad que en algunos casos puede llevar a responsabilidades desproporcionadas o, incluso, a censura.[43]

En el caso de las expresiones en Internet, las limitaciones que tienen problemas de vaguedad y de ambigüedad también pueden generar un efecto silenciador que lleve a la vulneración del derecho (ante la incertidumbre de lo que es válido hacer y lo que no lo es), y pueden "impactar especialmente en este uni-

[41] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 69.

[42] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 60.

[43] CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 70.



verso creciente de personas, cuya incorporación al debate público es una de las principales ventajas que ofrece internet como espacio de comunicación global".[44]

En segundo lugar (paso 2), las limitaciones deben cumplir tres condiciones: ser idóneas, necesarias y proporcionales. Decir que una limitación es necesaria para salvaguardar un objetivo democrático implica analizar si esa medida puede o no lograrse de la manera menos restrictiva (porque siempre debe optarse por la medida que limite menos la libertad de expresión). Decir que una limitación es idónea significa que resuelve efectivamente el problema en cuestión (no una que lo mantenga o agrave, por ejemplo). Decir que una limitación es proporcional atiende a la necesidad de no sacrificar excesivamente la libertad de expresión frente al bien que se protege; es decir, que hay una relación proporcional sobre el costo que la limitación implica para el derecho.[45]

La necesariedad de la medida no debe equipararse a una idea de utilidad ni de oportunidad,[46] sino que debe tratarse de una necesidad fuerte que hace que el objetivo no pueda protegerse por un medio menos restrictivo, al mismo tiempo que una vez que se reconoce que esto es así no debe limitarse más allá de lo indispensable (debe ser acotada al máximo de lo posible).[47]

La idoneidad funciona como un instrumento para evaluar el cumplimiento de la finalidad de la medida, que debe estar siempre limitada dentro del marco de

[44] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 58

[45] Corte I.D.H. Caso Herrera Ulloa vs. Costa Rica. Excepciones preliminares, Fondo Reparaciones y Costas. Sentencia de 2 de julio de 2004. Serie C, N.º 107, Párr. 121; Caso Gomes Lund y otros ("Guerrilha do Araguaia") Vs. Brasil. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 24 de noviembre de 2010; y Caso Claude Reyes y otros Vs. Chile. Fondo, Reparaciones y Costas. Sentencia de 19 de septiembre de 2006. Serie C No 151.

[46] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 46; Corte I.D.H., Caso Herrera Ulloa vs. Costa Rica. Sentencia de 2 de julio de 2004, Serie C No. 107, párr. 122; CIDH. Informe Anual 1994. Capítulo V: Informe sobre la Compatibilidad entre las Leyes de Desacato y la Convención Americana sobre Derechos Humanos. Título IV. OEA/Ser. L/V/II.88. doc. 9 rev. 17 de febrero de 1995.

[47] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párr. 83; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 85; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párrs. 121-122; Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985, Serie A No. 5, párr. 46.

interpretación democrático y de manera armónica con la libertad de expresión.[48]

La proporcionalidad obliga a quien establece la medida a intervenir en la menor medida posible con el ejercicio de la libertad de expresión, al tiempo de verificar que esa afectación efectivamente beneficia al interés protegido.[49] Para saber si esto es así, quien revise la medida debe evaluar el grado de afectación del derecho (grave, media, baja), la importancia del interés protegido (alta, media, baja) y el costo-beneficio de ese balance para ver si la restricción está justificada (en términos de proporcionalidad).[50]

En tercer lugar (paso 3), la evaluación del daño y de la medida debe ser siempre contextual, lo que significa que la ponderación no debe hacerse en abstracto sino partir de las circunstancias concretas del caso en cuestión.[51] Al analizar el contexto, también debe hacerse un “test de interés público” sobre la información relacionada con la expresión que se busca limitar (y así saber el grado de protección que ésta tiene). La información relacionada con el Estado, la gestión de gobierno, la transparencia, la rendición de cuentas de funcionarios públicos y otra de carácter similar alcanza este estándar de interés público, por lo que obtiene la protección reforzada que señalamos anteriormente.[52]

Finalmente, y para los casos en que se trate del derecho a la libertad de expresión en Internet, hay un cuarto paso que debe cumplirse que correspon-

[48] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177

[49] Corte I.D.H., Caso de Eduardo Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No.177, párr. 83; Corte I.D.H., Caso Palamara Iribarne. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 85; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 123; Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 de 13 de noviembre de 1985. Serie A No. 5, párr. 46; CIDH. Alegatos ante la Corte Interamericana en el caso Herrera Ulloa Vs. Costa Rica. Transcritos en: Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 101.1.B

[50] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párr. 84.

[51] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párr. 51; Corte I.D.H., Caso Tristán Donoso Vs. Panamá. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 27 de enero de 2009 Serie C No. 193, párr. 93.

[52] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párrs. 57 y 87; Corte I.D.H., Caso Claude Reyes y otros Vs. Chile. Sentencia de 19 de septiembre de 2006. Serie C No. 151, párrs. 84, 86 y 87; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 83; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 127. Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C. No. 107, párr. 106; CIDH (2009), Marco Jurídico Interamericano sobre la Libertad de Expresión, OEA, párr. 113



de a lo que la CIDH ha denominado la "perspectiva sistémica digital", y que significa que al evaluar la validez de una limitación a la libertad de expresión por Internet debe tomarse en cuenta el impacto que esa medida tiene en el funcionamiento de Internet en general, particularmente en cuanto a sus características fundamentales de ser una red descentralizada, libre y abierta.[53] Las limitaciones a la libertad de expresión en Internet afectan a toda la red, ya no solamente al ejercicio del derecho concreto, y por ello es importante pensar en las consecuencias que habilitar una limitación podría tener sobre el diseño propio de Internet.[54]

e. La prohibición de censura previa y de restricciones indirectas

El SIDH prohíbe tajantemente la censura previa con excepción de los espectáculos públicos en los que se proteja a la infancia y la adolescencia de acuerdo al artículo 13.4 de la CADH.[55] La Corte IDH ha sido enfática al decir que cuando existe una medida de censura previa se viola de *manera radical* el derecho a la libertad de expresión, y se afecta a la democracia en general.[56]

Es importante entender que las reglas de prohibición de censura son tanto para acciones directas como para medidas indirectas, sea que vayan dirigidas hacia los medios o insumos que requiere un medio de comunicación para poder difundir la información o hacia la forma en que ésta se difunde por Internet (por ejemplo con la remoción de enlaces o *links* o la moderación de contenidos).[57] [58]

[53] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 63.

[54] Centro de Estudios de Derecho, Justicia y Sociedad, Dejusticia. *El derecho a la libertad de expresión*, Colombia, 2017, p. 281.

[55] Corte I.D.H., Caso Kimel Vs. Argentina. Sentencia de 2 de mayo de 2008. Serie C No. 177, párr. 54; Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 79; Corte I.D.H., Caso Herrera Ulloa Vs. Costa Rica. Sentencia de 2 de julio de 2004. Serie C No. 107, párr. 120.

[56] Corte I.D.H., Caso Palamara Iribarne Vs. Chile. Sentencia de 22 de noviembre de 2005. Serie C No. 135, párr. 68; Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985. Serie A No. 5, párr. 54.

[57] Corte I.D.H., Caso "La Última Tentación de Cristo" (Olmedo Bustos y otros) Vs. Chile. Sentencia de 5 de febrero de 2001. Serie C No. 73; CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párr. 147.

[58] Corte I.D.H., Caso Ivcher Bronstein Vs. Perú. Sentencia del 6 de febrero de 2001. Serie C No. 74, párrs. 158 a 163.

El artículo 13.3 de la CADH señala que no debe restringirse este derecho por "vías o medios indirectos, tales como el abuso de controles oficiales o particulares de papel para periódicos, de frecuencias radioeléctricas, o de enseres y aparatos usados en la difusión de información o por cualesquiera otros medios encaminados a impedir la comunicación y la circulación de ideas y opiniones". Pero esas medidas no son excluyentes de otras que, en la actualidad y a la luz de las nuevas tecnologías, podrían constituirse también como medios de censura indirecta. Por esta razón la Corte IDH ha señalado expresamente que dicha enunciación no es taxativa y que deben evaluarse aquellos medios o vías indirectas que puedan tener esos efectos.[59]

Así como el derecho a expresarnos puede vulnerarse por distintos medios, también es importante entender que esto puede suceder con distintos sujetos: no sólo el Estado puede limitar la libertad de expresión, sino también los particulares. La CADH (vía el artículo 13.3) obliga a los Estados a proteger a las personas frente a los controles o interferencias de particulares que tengan como resultado limitar el derecho a expresarse libremente.[60]

Por eso los Estados parte de la CADH también pueden ser responsables por violar la Convención cuando permitan, fomenten o sean omisos a actuar frente a medidas de particulares que violen la libertad de expresión.[61] En dicha intersección (de lo público con lo privado), es particularmente relevante la obligación de no discriminación que todas las medidas deben respetar. No es válido que ninguna limitación fomente o perpetúe prejuicios o intolerancia frente a grupos vulnerabilizados, sea que esas medidas sean establecidas por los particulares o por el Estado.[62]

[59] Corte I.D.H., Caso Ríos y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 194, párr. 340; Corte I.D.H., Caso Perozo y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 195, párr. 367.

[60] Corte I.D.H., Caso Perozo y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 195, párr. 367; Corte I.D.H., Caso Ríos y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 194. párr. 240.

[61] Corte I.D.H., La Colegiación Obligatoria de Periodistas (Arts. 13 y 29 Convención Americana sobre Derechos Humanos). Opinión Consultiva OC-5/85 del 13 de noviembre de 1985. Serie A No. 5, párr. 48.

[62] CIDH. Informe Anual 1994. Capítulo V: Informe sobre la Compatibilidad entre las Leyes de Desacato y la Convención Americana sobre Derechos Humanos. Título III. OEA/Ser. L/V/II.88. doc. 9 rev. 17 de febrero de 1995.



f. Funcionarios públicos y la libertad de expresión

La Corte IDH ha señalado que además de las medidas administrativas y legislativas que pueden violar el derecho a expresarse libremente, los Estados pueden afectar este derecho al manifestarse públicamente a través de sus funcionarios públicos.

Si un discurso público incrementa o produce la vulnerabilidad de un grupo o persona, la libertad de expresión es violentada: si un gobierno se pronuncia en medios de comunicación de manera que intimide o restrinja la capacidad de ejercer el derecho, genera una situación de riesgo (que debería prevenir en primer lugar) para el derecho de ese grupo o persona.[63] La Corte ha realizado una evaluación sobre el poder y la desigualdad en la que algunos grupos se encuentran, para señalar que los funcionarios públicos deben reservarse de realizar expresiones que aumenten la "vulnerabilidad relativa" de los grupos que están en riesgo.[64] Además los Estados tienen una serie de deberes que deben de tomar en cuenta al momento de realizar manifestaciones, tales como:[65]

- Deber de pronunciarse en ciertos casos, en cumplimiento de sus funciones constitucionales y legales, sobre asuntos de interés público.
- Deber especial de constatación razonable de los hechos que fundamentan sus pronunciamientos.
- Deber de asegurarse de que sus pronunciamientos no constituyan violaciones a los derechos humanos.
- Deber de asegurarse de que sus pronunciamientos no constituyan una injerencia arbitraria, directa o indirecta, en los derechos de quienes contribuyen a la deliberación pública mediante la expresión y difusión de su pensamiento.

[63] Corte I.D.H. Caso Perozo y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 195, párr. 161; Corte I.D.H.,Caso Ríos y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 194, párr. 149.

[64] Corte I.D.H., Caso Ríos y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 194, párr. 145; Corte I.D.H., Caso Perozo y otros Vs. Venezuela. Excepciones Preliminares, Fondo, Reparaciones y Costas. Sentencia de 28 de enero de 2009. Serie C No. 195, párr. 157.

[65] *Cfr.* en CIDH, Relatoría Especial para la Libertad de Expresión. Marco jurídico interamericano sobre la libertad de expresión. OEA/Ser.L/V/II. CIDH/RELE/INF.2/09. 30 diciembre 2009, párrs. 201 - 205.

- Deber de asegurarse de que sus pronunciamientos no interfieran sobre la independencia y autonomía de las autoridades judiciales.

g. La libertad de expresión e Internet

Las características intrínsecas de Internet lo han vuelto una verdadera herramienta facilitadora y habilitadora de otros derechos, razón por la cuál ha sido reconocido en algunos países como un derecho humano configurado de manera autónoma.[66]

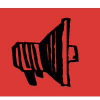
Por su importancia, en el 2011 el Relator Especial de las Naciones Unidas (ONU) para la Libertad de Opinión y de Expresión, la Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), la Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión y la Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP), elaboraron una declaración conjunta sobre Libertad de Expresión e Internet en la que en su punto 1 c) enfatizaron que:

> *"Los enfoques de reglamentación desarrollados para otros medios de comunicación —como telefonía o radio y televisión— no pueden transferirse sin más a Internet, sino que deben ser diseñados específicamente para este medio, atendiendo a sus particularidades".*[67]

La lógica detrás de esta regla es que no es posible tratar de la misma manera a Internet que como se trata a otros medios de comunicación porque es una forma de comunicación con particularidades que requieren un trato específico para mantenerlo como un espacio libre y abierto. La arquitectura de Internet tiene, por ejemplo, el elemento de la neutralidad de la red, que ha sido definida en términos generales como la facilitación al "acceso y la difusión de contenidos, aplicaciones y servicios de manera libre y sin distinción alguna. Al mismo tiempo, la inexistencia de barreras desproporcionadas de

[66] México por ejemplo, con el derecho de acceso a las tecnologías de la información en su artículo 6o constitucional.

[67] ONU. OSCE. OEA. CADHP. Declaración conjunta sobre Libertad de Expresión e Internet. 2011. Disponible en: http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=849&lID=2



entrada para ofrecer nuevos servicios y aplicaciones en Internet constituye un claro incentivo para la creatividad, la innovación y la competencia".[68]

En ese mismo sentido, la CIDH en su informe sobre Libertad de Expresión e Internet reiteró que:

> *"Internet se ha desarrollado a partir de determinados principios de diseño, cuya aplicación ha propiciado y permitido que el ambiente en línea sea un espacio descentralizado, abierto y neutral. Es importante que cualquier regulación que se produzca [...] mantenga las características básicas del entorno original, potenciando su capacidad democratizadora e impulsando el acceso universal y sin discriminación".* [69]

Además:

> *"todas las medidas que puedan de una u otra manera afectar el acceso y uso de Internet deben interpretarse a la luz de la primacía del derecho a la libertad de expresión, sobre todo en lo que respecta a los discursos especialmente protegidos en los términos del artículo 13 de la Convención Americana".*[70]

Así, es claro que la protección del principio de la neutralidad de la red es fundamental para garantizar la pluralidad y la diversidad de la información que fluye a través de Internet. También es claro que la arquitectura de esta "red de redes" es fundamental para equilibrar y potenciar un debate público democrático, que sea de carácter inclusivo y plural (que tenga en su base un "pluralismo informativo").[71]

[68] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 27.; Consejo de Europa. Comité de Ministros. Declaration of the Committee of Ministers on network neutrality. 29 de septiembre de 2010. Punto 3; Belli, Luca. Council of Europe Multi-Stakeholder Dialogue on Network Neutrality and Human Rights. Outcome Paper. CDMSI(2013)Misc 18. 3 a 6 de diciembre de 2013. Párr. 16-17.

[69] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 11.

[70] Ibid., párr. 14.

[71] Corte IDH. Caso Kimel Vs. Argentina. Fondo, Reparaciones y Costas. Sentencia de 2 de mayo de 2008. Serie C No. 177. Párr. 57; Corte IDH. Caso Fontevecchia y D`Amico Vs. Argentina. Fondo, Reparaciones y Costas. Sentencia de 29 de noviembre de 2011. Serie C No. 238. Párr. 45.

En este sentido, toda la legislación relacionada con la libertad de expresión y el Internet debe entonces tomar en cuenta sus características y pensar que los controles y medidas que constituyan límites a la libertad de expresarse y a la de acceder libremente a Internet, deben de cumplir los estándares que establece el SIDH que desarrollados previamente, además de aquellos que resulten de las especificidades estructurales de Internet.



B. LA RESPONSABILIDAD DE LOS INTERMEDIARIOS EN INTERNET FRENTE A EXPRESIONES DE TERCEROS.

En este apartado, se abordará el rol que juegan los intermediarios (primordialmente privados) en Internet y la manera en la que gestionan las expresiones de terceros que son publicadas, alojadas, transmitidas o enlazadas a través de dichos servicios.

Las discusiones en torno a la responsabilidad de los intermediarios se han intensificado en Latinoamérica en los últimos años, a partir de distintas propuestas legislativas en la región, de la adopción de tratados de libre comercio y de los marcos regulativos europeos en la materia. Esta discusión es fundamental por su centralidad para el ejercicio de derechos como la libertad de expresión, pero también para un conjunto más amplio de derechos que se ejercen en Internet (relacionados fuertemente con el desarrollo de distintos servicios en línea).[72]

a. El rol de los intermediarios en Internet

Los intermediarios son actores privados que brindan una serie de servicios tales como el acceso y la interconexión; la transmisión, procesamiento y encaminamiento del tráfico; el alojamiento de material publicado por terceros y el acceso a éste; la referencia a contenidos o la búsqueda de materiales en la red; la realización de transacciones financieras; y la conexión entre usuarios a través de plataformas de redes sociales (entre otros).[73] La CIDH señala que aunque hay distintas formas de clasificarlos, las más relevantes son:[74]

[72] Del Campo, Agustina; Schatzky, Morena; Hernández, Laura; Lara, Juan Carlos. *Mirando al Sur. Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios en Internet*, Al Sur, Abril 2021, pp. 4-8.

[73] Naciones Unidas. Asamblea General. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Frank La Rue. A/HRC/17/27. 16 de mayo de 2011. Párr. 38. Disponible para consulta en: http://ap.ohchr.org/documents/dpage_s.aspx?m=8

[74] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 91.

- Proveedores de servicios de Internet (PSI)
- Proveedores de alojamiento de sitios Web
- Plataformas de redes sociales
- Motores de búsqueda

En gran medida, los intermediarios son responsables de impulsar el impacto social de la libertad de expresión, razón por la cual muchas veces sus acciones quieren ser controladas a través de la imposición de responsabilidades sobre ellos y sobre la posición que ocupan y el rol que cumplen. Sin exagerar, los intermediarios se han erigido como los puntos a través de los que es (técnicamente) posible ejercer el control de los contenidos en Internet.[75]

b. El principio de la no responsabilidad de intermediarios

La importancia de evitar afectaciones a la libertad de expresión en su dimensión individual y -particularmente- social ha sido reconocida a través de la Declaración Conjunta sobre Libertad de Expresión e Internet, establecida por el Relator Especial de las Naciones Unidas (ONU) para la Libertad de Opinión y de Expresión, la Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), la Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión y la Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP):[76]

> *"2. Responsabilidad de intermediarios*
>
> *a. Ninguna persona que ofrezca únicamente servicios técnicos de Internet como acceso, búsquedas o conservación de información en la memoria caché deberá ser responsable por contenidos generados por terceros y que se difundan a través de estos servicios, siempre que no intervenga específicamente en dichos*

[75] Naciones Unidas. Asamblea General. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Frank La Rue. A/HRC/17/27. 16 de mayo de 2011. Párr. 74. Disponible para consulta en: http://ap.ohchr.org/documents/dpage_s.aspx?m=8

[76] ONU. OSCE. OEA. CADHP. Declaración conjunta sobre Libertad de Expresión e Internet. 2011. Disponible en: http://www.oas.org/es/cidh/expresion/showarticle.asp?artID=849&lID=2

contenidos ni se niegue a cumplir una orden judicial que exija su eliminación cuando esté en condiciones de hacerlo ("principio de mera transmisión").

b. Debe considerarse la posibilidad de proteger completamente a otros intermediarios, incluidos los mencionados en el preámbulo, respecto de cualquier responsabilidad por los contenidos generados por terceros en las mismas condiciones establecidas en el párrafo 2(a). Como mínimo, no se debería exigir a los intermediarios que controlen el contenido generado por usuarios y no deberían estar sujetos a normas extrajudiciales sobre cancelación de contenidos que no ofrezcan suficiente protección para la libertad de expresión (como sucede con muchas de las normas sobre "notificación y retirada" que se aplican actualmente)."

El énfasis que la Declaración Conjunta da al rol de la no responsabilidad de intermediarios no es mera ocurrencia, sino que refleja que su lugar privilegiado para ejercer control sobre los contenidos que circulan en Internet los vuelve un blanco buscado a menudo por los gobiernos para interferir en el flujo de la información. La presión resulta de que es más fácil identificar y constreñir a estos actores que a los responsables directos de la expresión que se busca inhibir o controlar. [77]

Sin embargo, la propia declaración refleja el consenso internacional que existe sobre el rechazo a los modelos de responsabilidad objetiva de los intermediarios (que implica responsabilizarlos por contenidos ilegítimos o ilegales generados por terceros).[78] Una de las razones principales de este consenso es la dificultad de revisar todos los contenidos que circulan en, por ejemplo, la plataforma de un intermediario, y de la carga que implica presumir que evitar el daño potencial de un tercero es una acción que está dentro del control limitado que los intermediarios en verdad poseen. El consenso respalda que los intermediarios no deben estar sujetos legalmente a obligaciones de supervisión de los

[77] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 93.
[78] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 95.



contenidos generados por los usuarios con el fin de detener y filtrar expresiones ilícitas.[79]

La CIDH propone una analogía bastante útil para explicar el impacto antidemocrático que generaría responsabilizar a los intermediarios de manera objetiva por la circulación de información generada por terceros: responsabilizar a un intermediario en este sentido, en el contexto de una red abierta, plural, universalmente accesible y expansiva, sería como responsabilizar a las compañías de teléfono por las amenazas telefónicas que una persona hace a otra, causándole incertidumbre o algún otro tipo de daño.[80]

El Relator especial para la libertad de expresión de la ONU sostuvo, en el mismo sentido, que responsabilizar a intermediarios por el contenido difundido o creado por sus usuarios lesiona gravemente el disfrute del derecho a la libertad de expresión porque produce un tipo de censura privada,[81] que

[79] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 96.; Relator Especial de las Naciones Unidas (ONU) sobre la Promoción y Protección del derecho a la Libertad de Opinión y de Expresión, Representante para la Libertad de los Medios de Comunicación de la Organización para la Seguridad y la Cooperación en Europa (OSCE), Relatora Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión, y Relatora Especial sobre Libertad de Expresión y Acceso a la Información de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP). 1 de junio de 2011. Declaración conjunta sobre libertad de expresión e Internet. Punto 2 (b); Tribunal de Justicia de la Unión Europea. Scarlet Extended SA v. Société belge des auteurs, compositeurs et éditeurs SCRL (SABAM). C-70/10. Sentencia de 24 de noviembre de 2011. Párr. 49-53; Tribunal de Justicia de la Unión Europea. Belgische Vereniging van Auteurs, Componisten en Uitgevers CVBA (SABAM) v. Netlog NV. C-360/10.Sentencia de 16 de febrero de 2012. Párr. 47- 51.

[80] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 97.

[81] Existe cierto consenso de que las iniciativas que se han presentado en los últimos años en distintos países (particularmente en Europa) son altamente problemáticas para las preocupaciones aquí señaladas y pensando dentro del marco interamericano de derechos humanos. Así, un análisis reciente señala que: "Las iniciativas regulatorias de los últimos años fueron criticadas, principalmente, por sus efectos adversos en los derechos humanos, con particular atención en el derecho a la libertad de expresión. Sucede que las amenazas de responsabilidad, con importantes multas pecuniarias –o, peor aún, penas de prisión, como es el caso de la ley australiana–, sumado a la presión de resolver en plazos extremadamente breves, generan un incentivo de exceso de remoción de contenido conocido como "censura privada". Ante estas presiones, el temor es que las plataformas eliminen contenido presunta o manifiestamente ilegal y, en muchos casos, completamente legal, violando la protección del derecho a la libertad de expresión reconocido en instrumentos internacionales". Del Campo, Agustina; Schatzky, Morena; Hernández, Laura; Lara, Juan Carlos. Mirando al Sur. Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios en Internet, Al Sur, Abril 2021, p. 30.

surge como una respuesta de autoprotección de los intermediarios (para evitar ser sancionados) que es excesivamente amplia, poco transparente y sin las debidas garantías procesales.[82]

Por estas razones, la CIDH sostiene que la responsabilidad de intermediarios por expresiones de un tercero que resulten ilícitas sólo debe proceder cuando se ordene por una autoridad judicial que opere con suficientes garantías de independencia, autonomía e imparcialidad, y que sea capaz de evaluar los derechos que están en juego para ofrecer las garantías necesarias al usuario (por lo que las resoluciones o recomendaciones de mecanismos u organismos de naturaleza administrativa quedarían excluidos en principio).[83]

c. La Sección 230 de la Communications and Decency Act de los Estados Unidos de América

El principio de la no responsabilidad de intermediarios tiene su origen en la legislación de los Estados Unidos de América (EUA). A raíz de la influencia que la regulación de EUA ha tenido en el desarrollo de Internet, uno de los instrumentos jurídicos más relevantes en relación a la responsabilidad de intermediarios es el de la Sección 230 de la Ley de Decencia en las Comunicaciones, que fue agregada a la *Ley de Telecomunicaciones* en este país.[84]

Esta sección, por un lado, reconoce el principio de la no responsabilidad de intermediarios por las expresiones de sus usuarios. Sin embargo, también aborda la ausencia de responsabilidad por las acciones de moderación

[82] ONU. Relator Especial de las Naciones Unidas para la libertad de expresión, A/HRC/17/27, párr. 40.

[83] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 106.

[84] Esta disposición se articula, desde luego, con la Primera Enmienda de la Constitución norteamericana, la cual tiene un valor central en el esquema constitucional estadounidense y que en algunas ocasiones incluso ha llegado a ser considerada como “absoluta”. El alcance de la Primera Enmienda se da en principio para el Estado pero sin duda también incluye a las empresas y, en el caso que nos ocupa, a los intermediarios de Internet. La sección 230 establece el régimen de moderación de contenidos en Internet de manera articulada con la Primera Enmienda desde comienzos de los años 90. Al respecto ver: Del Campo, Agustina; Schatzky, Morena; Hernández, Laura; Lara, Juan Carlos. *Mirando al Sur. Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios en Internet*, Al Sur, Abril 2021, p. 17.



de contenidos voluntariamente adoptadas por los intermediarios (regla del "buen samaritano", tal como explicamos más adelante):

1. "Ningún proveedor o usuario de un servicio de cómputo interactivo deberá ser tratado como editor o emisor de cualquier información proveída por otro proveedor de servicios.
2. Ningún proveedor o usuario de un servicio de cómputo interactivo será responsable legalmente de:

   a) Ninguna acción voluntaria de buena fe que restrinja el acceso o la disponibilidad del material que el proveedor o usuario considere ser obsceno, vulgar, lascivo o excesivamente violento, abusivo o de otra manera indeseable, ya sea material protegido constitucionalmente o no.

   b) Ninguna acción tomada para habilitar o hacer disponible a proveedores de contenido informativo u otros las medidas técnicas para restringir el acceso a material descrito en el párrafo 1".[85]

i. El origen de la sección 230.

La Primera Enmienda a la Constitución de EUA posee abundante jurisprudencia relativa a la diferencia entre el sujeto que distribuye expresiones (como una televisora, una imprenta o un programa de radio) y la tercera persona que emite dichas expresiones.[86] La jurisprudencia señala que un distribuidor no será responsable legalmente de lo que haya expresado un tercero siempre y cuando no haya sabido o no hubiera debido tener conocimiento sobre el contenido que genera dicha responsabilidad.[87]

Esta doctrina surgió y fue pensada, desde luego, en una época donde el Internet no existía y las discusiones sobre el derecho a la libertad de expresión

[85] Traducción propia. Consulta en el idioma original en: https://www.law.cornell.edu/uscode/text/47/230#f_3.

[86] La protección reforzada de la sección 230 ha llevado a que dentro de la academia se considere que el estándar establecido en ella supera los alcances de la Primera Enmienda, convirtiéndose en un a ley o estatuto que potencia la libertad de expresión (*speech-enhancing statute*), por alcanzar "no solo contenido difamatorio sino cualquier denuncia basada en el contenido de terceros". Del Campo, Agustina; Schatzky, Morena; Hernández, Laura; Lara, Juan Carlos. Mirando al Sur. Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios en Internet, Al Sur, Abril 2021, p. 18, citando a: Goldman, Eric, "Why section 230 is Better Than The First Amendment", Notre Dame Law Review Reflection, 2019.

[87] Cfr. Koseff, Jeff. "The Twenty Six Words That Created The Internet". Cornell University Press. New York. 2019. p. 11-35.

se centraban en el papel que tenían los medios masivos de comunicación. Por lo mismo, encontró complicaciones cuando las personas usuarias empezaron a usar internet. La creación de la sección 230 de la CDA se explica en gran parte a partir de dos casos de la Corte Suprema de Justicia de este país: *Cubby Inc vs CompuServe* y *Stratton Oakmont Inc v Prodigy Services Co*.

El caso *Cubby* se refiere a una situación en la que se realizó una supuesta calumnia hacia la empresa *Cubby*, por medio de un boletín llamado "Rumorville", en un foro de Internet llamado *Compuserve*. Compuserve no realizaba ningún "control editorial" sobre los contenidos antes de que se publicaran. La Corte decidió que como *Compuserve* no hacía una revisión activa de su sitio, la naturaleza de la plataforma era de un distribuidor y no de editor, por lo que no debía ser sujeto a responsabilidades de ese estilo.[88]

Sin embargo, en el caso Stratton, donde una persona publicó comentarios "difamatorios" contra una firma de corretajes reconocida en un foro de la empresa *Prodigy*, la Corte decidió de manera muy controversial que la empresa sí era responsable legalmente por fungir como editora de dichos foros. La Corte argumentó que el trabajo de la empresa debía considerarse como trabajo editorial ya que ella misma reconocía que hacía moderación de sus contenidos y que eliminaba de manera activa algunas publicaciones en su foro.

La última decisión fue altamente mediática, y recibió la crítica de varias personas expertas, quienes mostraron su preocupación por dicho incidente, ya que en su opinión la decisión abría las puertas a la arbitrariedad para determinar la responsabilidad legal de las plataformas. Muchas personas consideraron que la moderación de *Prodigy* no suponía un trabajo editorial y que el problema principal de la decisión judicial era que, dada la ambigüedad que generaba, se constituía como un precedente que incentivaría a los intermediarios a moderar y reducir contenido malicioso, para no correr el riesgo de ser imputados con esta responsabilidad legal.[89]

Las preocupaciones que se presentaron a partir de las decisiones de la Corte Suprema de Justicia se tradujeron en una discusión en sede legislativa, en el

[88] Cubby, Inc. v. CompuServe, Inc. 776 F. Supp. 135 (1991). Disponible en: https://law.justia.com/cases/federal/district-courts/FSupp/776/135/2340509/.

[89] Kosseff Jeff. "The Twenty Six Words…", op. cit., p. 53.



Congreso de los EUA. Los congresistas que impulsaron esta normativa estaban buscando una manera de generar los incentivos apropiados para moderar contenido y lograr el desarrollo de la industria. Al analizar el dilema jurídico, llegaron a la conclusión de que la sobre-regulación de los intermediarios obstruía de manera grave la creación y desarrollo de nuevos servicios en línea.[90]

Tras la discusión, se estableció un apartado que establece, en la sección 230 c, dos puntos principales:

1. La responsabilidad legal derivada de los contenidos de terceros publicados en plataformas interactivas recae sobre el proveedor de esta información y no en la plataforma.
2. Las medidas unilaterales de moderación realizadas por las plataformas de "buena fe" son permitidas y no causan responsabilidad legal.

¿Qué se busca al establecer ambos principios? Generar incentivos positivos para los intermediarios. Lo que el Congreso buscó al incorporar estos apartados fue no desincentivar la expansión de los intermediarios en Internet, en especial al incentivar que quienes puedan bloquear contenido violento u ofensivo pudieran hacerlo en el marco del principio del "buen samaritano" señalado previamente.[91] Posteriormente, en sede judicial, la Corte suprema reconoció que los dos puntos más importantes de esta disposición eran: i) la protección de las medidas de buena fe que remueven contenido "indecente" y ii) la protección a la libertad de expresión.[92]

ii. El principio de no responsabilidad de intermediarios en la sección 230

Como mencionamos más arriba, la distinción entre el sujeto editor y el sujeto autor o emisor de un mensaje había sido uno de los temas más importantes para la jurisprudencia relacionada con la Primera Enmienda de la Constitución de EUA, la que es considerada cotidianamente como la piedra angular de la libertad de expresión en este país. El caso *Zeran* mantuvo este enfoque al establecer que las plataformas no serían legalmente responsa-

[90] Id. p.60
[91] Klonick, Kate. "The new governors: The people, rules, and processes governing online speech". *Harvard Law Review*, 131, 2018, p. 1605.
[92] Id. p. 1608, en el caso *Zeran vs America*.

bles incluso si tenían conocimiento de la producción de conocimiento objetable de un tercero.[93]

La lógica detrás del fallo fue que sólo de esta forma los intermediarios no tendrían un incentivo perverso de remover cualquier tipo de mensaje que les era notificado o que era encontrado de manera circunstancial. De otra forma, como sucedería con un sistema de responsabilidad objetiva, las plataformas tendrían la presión de establecer sistemas de vigilancia que resultarían en censura de contenidos, aunque esta fuera "colateral".

La censura colateral ocurre cuando una entidad privada puede controlar el discurso de sus usuarios a través de sistemas de moderación.[94] Si la regulación permite que los intermediarios sean responsables legalmente por el contenido de terceros, estas empresas toman un rol activo de censurar cualquier expresión que pueda acarrear el mínimo riesgo de una demanda.[95] Es decir, que ante el riesgo de responsabilidades y ante la duda frente a ciertas expresiones, las corporaciones sólo obtienen incentivos de vigilancia y control, y no así de proteger y garantizar la libertad de expresión.

Precisamente ante este tipo de situaciones, la sección 230 protege la libertad de expresión de las personas usuarias al evitar esos incentivos perversos. Esta visión fue compartida por la Corte Suprema de Justicia en el caso *Reno vs ACLU*, donde se discutió la constitucionalidad de varias disposiciones del *Acta de Decencia en las Comunicaciones*, en lo relativo a la protección a menores frente a contenido "pornográfico o indecente". La Corte Suprema estableció que Internet es un medio radicalmente distinto a los medios masivos de comunicación tradicionales, tal como la radio o la televisión, y que la diferencia radica en que mientras que estos últimos limitan la entrada a ciertos creadores de contenido, el Internet habilita a que cualquier persona usuaria pueda publicar y compartir contenido en línea.

Dado que Internet cuenta con una arquitectura libre y abierta, que es diametralmente distinta a la de los medios tradicionales (restringida y limitada),

[93] Kosseff Jeff. "The Twenty Six Words..." *op. cit.*, p. 86-102.

[94] Ver Michael Meyerson Meyerson, Michael I. "Authors, Editors, and Uncommon Carriers: Identifying the 'Speaker' Within the New Media" (1995). *Notre Dame Law Review*, Vol. 71, No. 1, p. 79, 1995, Available at SSRN: https://ssrn.com/abstract=1327090.

[95] Idem.



el potencial igualador y maximizador de la libertad de expresión radica en que dicha arquitectura se mantenga así. En este caso, el Juez Stevens reconoció que para que Internet mantuviera su crecimiento (en sintonía con su efecto potenciador de la libertad de expresión), era necesaria la mínima intervención a las expresiones en general, por lo que consideró que las cláusulas y sanciones relativas al concepto de "indecencia" eran inconstitucionales, para luego dejar intacta el resto de la sección 230.

La sección 230 en su estado actual se enfoca fuertemente hacia facilitar el auge económico y tecnológico de las empresas estadounidenses. Al hacerlo, da un margen amplio que resulta en beneficio del derecho a la libertad de expresión porque se preocupa por evitar la censura de expresiones lícitas, que podrían realizar los intermediarios que buscaran evitar riesgos innecesarios (frente a un sistema distinto de responsabilidad objetiva).[96] Por estas razones, la primera parte de la sección 230 se materializa como una de las bases principales del Internet moderno, que es indispensable para el ejercicio de la libertad de expresión en línea, independientemente de todos los debates relacionados frente a las obligaciones que algunos intermediarios, principalmente privados, tienen frente a la libertad de expresión y a otros derechos que se ejercen en el entorno digital.

iii. La no responsabilidad por medidas unilaterales respecto de la moderación de contenidos

Con el objetivo de favorecer el desarrollo de Internet, así como incentivar que las plataformas privadas pudieran tomar acciones para remover expresiones abusivas en línea, además de la regla de la no responsabilidad de los intermediarios por las expresiones de terceros, la Sección 230 incorpora una disposición relacionada con las medidas unilaterales de moderación de contenidos. En este caso, el apartado 2 de la Sección 230 excluye de responsabilidades legales a las medidas unilaterales de moderación que sean de buena fe y sobre contenidos que puedan considerarse objetables o indeseables (como mencionamos anteriormente, la regla del "buen samaritano").

[96] Keller, Daphne. "El "derecho al olvido" de Europa en América Latina", en Del Campo, Agustina (coord.). *Hacia una Internet libre de censura II: perspectivas en América Latina*. Universidad de Palermo, Facultad de Derecho, Centro de Estudios en Libertad de Expresión y Acceso a la Información, Buenos Aires, 2017, p. 180. Disponible en: https://www.palermo.edu/cele/pdf/investigaciones/Hacia_una_internet_libre_de_censura_II.pdf

La libertad de gestionar los contenidos generados por terceros otorgada por la regla del "buen samaritano" permite a los intermediarios contar con la flexibilidad necesaria para asegurar que su servicio o plataforma sea usable y atractivo para la mayoría de las personas, que sea viable económicamente, especialmente cuando el intermediario posee un modelo de negocios basado en la publicidad y en general, para poder remover contenido considerado perjudicial sin temor a represalias legales.

De esta manera, la regla del "buen samaritano" incorporada en la sección 230 presume generar beneficios al interés público y a las propios intermediarios al pretender generar un balance entre la ausencia de incentivos para la censura garantizada por la no responsabilidad de intermediarios por el contenido generado por terceros y la presencia de incentivos para actuar frente a contenidos "abusivos" publicados, alojados o enlazados en las plataformas o servicios ofrecidos por dichos intermediarios.

Así, la Sección 230 evita que los intermediarios puedan ser considerados jurídicamente como "editores", y les sean aplicables las responsabilidades derivadas de dicho carácter, por el solo hecho de llevar a cabo acciones de moderación de contenidos (siempre y cuando dicha moderación se realice de buena fe).[97]

d. El principio de no responsabilidad de intermediarios en tratados comerciales

El principio de la no responsabilidad de intermediarios ha comenzado a incluirse en tratados comerciales como el Tratado de México, Estados Unidos de América y Canadá, conocido como T-MEC, que sustituye al Tratado de Libre Comercio de América del Norte (TLCAN), y que a diferencia de este último incluye en su capítulo 19 una serie de disposiciones y condiciones que deben cumplir los países parte del acuerdo en torno al "Comercio Digital".[98]

[97] Gillespie, Tarleton. "Custodians of the Internet". Yale University Press. Estados Unidos. 2018. p.30-31

[98] La falta de regulación de distintos países en Latinoamérica hace que uno de los marcos normativos en materia de intermediarios se traiga "desde afuera" con las disposiciones sobre propiedad intelectual y el comercio digital de los tratados de libre comercio. En el caso de tratados con los Estados Unidos de América, lo más común es encontrar regulaciones que refieren a la Digital Mi-



En el apartado 17 de dicho capítulo, se reconoce, en términos similares a la Sección 230 descrita, el principio de que la no responsabilidad de intermediarios respecto al contenido generado por terceros que estas plataformas alojan o procesan, en los siguientes términos:

> *Ninguna Parte adoptará o mantendrá medidas que traten a un proveedor o usuario de un servicio informático interactivo como proveedor de contenido de información para determinar la responsabilidad por daños relacionados con la información almacenada, procesada, transmitida, distribuida o puesta a disposición por el servicio, excepto en la medida en que el proveedor o usuario, en su totalidad o en parte, haya creado o desarrollado la información.*

Los tratados de libre comercio han incorporado intereses y disposiciones que están en tensión no solo con las normativas locales de distintos países sino con los estándares interamericanos. En parte por esta razón, tanto la sociedad civil como la academia especializada en el ejercicio de los derechos humanos en el entorno digital han realizado campañas y recurrido a medidas judiciales que le hacen frente a la implementación de parte de las disposiciones de los tratados comerciales que ponen en riesgo la libertad de expresión en Internet y que se relacionan con el principio de la no responsabilidad de intermediarios.[99]

llennium Copyright Act (DMCA) de este país, que incluye la figura de la notificación y retiro o “Notice and Takedown”. Lo cierto es que aunque existen diferencias entre la forma en que los distintos países en Latinoamérica recogen estas disposiciones y no hay uniformidad, muchas de estas reglas son coincidentes. Para un análisis completo de este punto en la mayor parte de la región ver: Del Campo, Agustina, et. al. Mirando al Sur... op. cit. 72., pp. 9-11. Para ver cómo el mecanismo establecido en la DMCA funciona como una excepción a la sección 230 en términos de la ley de derechos de autor (que ha sido fuertemente criticada aún con su alcance limitado), revisar en específico pp. 19-20.

[99] Del Campo, Agustina, et. al. “Mirando al Sur...”. *op. cit.*, pp. 10-11.

C. LA MODERACIÓN DE CONTENIDOS

Como fue desarrollado previamente, la moderación de contenidos, entendida como la “práctica organizada de revisión de contenidos generados por usuarios publicados en páginas de internet, redes sociales u otras plataformas”,[100] ha sido considerada esencial para el funcionamiento de las plataformas en Internet. Para expertos como Tarleton Gillespie, una plataforma no solo sería poco funcional sin la moderación, sino que la moderación es un elemento indispensable para que exista una plataforma.[101]

Sin embargo, como se desarrollará en este apartado, la moderación de contenidos en la práctica representa desafíos complejos que han suscitado cuestionamientos a las premisas asumidas por disposiciones como la Sección 230 que explicamos arriba.

Un ejemplo de lo anterior es el objetivo de mantener el flujo informativo en Internet. ¿Cómo debe entenderse la necesidad de excluir contenidos atroces y asegurar el cumplimiento de este objetivo?[102] En muchos casos, la tensión entre ambos es inevitable y lleva a conflictos o dilemas en la moderación, que sólo pueden ser sorteados haciendo un balance de los intereses en juego.

Cuando problematizamos de esta forma la moderación de contenidos, es más fácil comprender las dos posturas más extremas (y opuestas) en el debate sobre los *qué’s* y el *cómo* de la moderación: aquellas personas que creen que las plataformas han sido muy permisivas con contenidos atroces (como la pornografía infantil, el acoso en línea o el discurso de odio que incita al a violencia) y aquellas que señalan que las platadormas se han excedido en sus facultades y han intervenido demasiado en la discusión pública.[103]

[100] Roberts, Sarah. “Behind The Screen: Content Moderation in the Shadows of Social Media”. Yale University Press. Estados Unidos. 2019. P. 33.
[101] Id. p. 21.
[102] Gillespie, Tarleton. “Custodians of the…”. *op. cit.*. p. 10.
[103] Idem, p. 11.

Para entenderlas mejor, es necesario comprender de forma más completa los incentivos que las plataformas tienen al moderar, las reglas de moderación, los procedimientos de moderación que se llevan a cabo y las consecuencias que tiene la moderación de contenidos en la libertad de expresión.

a. Objetivos y justificaciones de la moderación de contenidos

Todas las plataformas moderan. Aunque muchas de ellas evitan que la moderación sea muy notoria, es inevitable. Para otras, por sus características o por el servicio que proveen, el sistema de moderación es una de las principales características de su negocio.

Algunas de las principales razones para moderar son:

- **La imagen corporativa:**

  Si bien es cierto que muchas plataformas crean reglas y sistemas de moderación por una cuestión de responsabilidad social, es indiscutible que muchas otras lo hacen para que se adecue a su identidad corporativa.[104]

  No todas las plataformas son un foro abierto de discusión descentralizada y sobre cualquier temática. Por ejemplo, mientras que Facebook tiene la misión de hacer que el mundo esté más conectado y abierto, para generar un espacio donde puedan convivir amigos y familia, encontrar comunidades y crecer negocios,[105] Reddit tiene el propósito de albergar pequeñas comunidades para cualquier tema que le apasione a sus usuarios.

- **Las razones económicas:**

  Los incentivos económicos son quizás la razón más importante del actuar de muchas plataformas. Su objetivo principal atiende a un modelo

[104] Klonick, Kate. "The New Governors..." *op. cit.*, p. 1625.
[105] Meta. "Company Info". Consultado el 6 de diciembre de 2021. Disponible en: https://about.fb.com/company-info/.

de negocios que busca maximizar las ganancias posibles a partir de hacer que el usuario permanezca en su plataforma y así incrementar sus ganancias en publicidad.[106]

Un caso que permite observar esta lógica es el de la política contra el discurso de odio que incita a la violencia que fue llevada a cabo por Twitter (y que también puede verse como coincidente con un interés social de fortalecer el debate público), que se generó a partir de la reacción negativa de usuarios ante las amenazas, acoso selectivo y violencia contra feministas y periodistas en el controversial caso de *Gamergate*.[107]

b. Las reglas de moderación

Para llevar a cabo una moderación efectiva, las plataformas requieren establecer reglas claras y principios orientadores que les permitan a sus usuarias entender cuáles son las normas comunitarias con las que la plataforma generará su entorno de comunicación, así como para ordenar los sistemas de moderación y a las personas empleadas que se encargarán de las tareas de moderación de contenidos. Se trata, entonces, de las reglas del juego con las que una plataforma se relaciona con el público que participe en ella, por lo que deberán ser públicas y claras para todas las personas participantes.

La publicidad y claridad de las reglas de moderación son fundamentales para que los intermediarios actúen dentro de un marco transparente en el que puedan ser sujetas a rendir cuentas. También son indispensables para limitar la arbitrariedad con la que dichas plataformas pueden actuar (y muchas veces efectivamente actúan).

Existen dos textos fundamentales para las plataformas que involucran la publicación de contenido de terceros y que posibilitan el escrutinio público sobre sus acciones: las normas comunitarias y los términos y condiciones.

Los términos y condiciones son un contrato por el cual se establecen las

[106] Klonick, Kate. "The New Governors..." *op. cit.*, p. 1627.
[107] Idem, p. 1629.



obligaciones del usuario y la plataforma; se señalan los métodos de resolución de controversias, los contenidos que son adecuados y también hacen referencia a las responsabilidades legales, propiedad intelectual o cualquier referencia que pueda evitar un litigio.[108]

Las normas comunitarias suelen establecerse como documentos dirigidos a las personas usuarias, que se redactan en un lenguaje claro para ser entendibles por todas ellas. Ahí se especifican las conductas esperadas y las que no tienen cabida dentro de su red social;[109] se detallan de manera más específica sus "valores", su visión y la clase de interacciones que propician en sus espacios con sus usuarios.

Si bien cada plataforma es distinta, normamente se establecen mínimos de prohibiciones en contenidos como *spam*, pornografía explícita, discurso de odio que incite a la violencia, acoso o contenido ilegal de acuerdo a la normativa del lugar de acceso. Más adelante se abordarán los riesgos y consecuencias que puede haber en una definición defectuosa de estas prohibiciones.

También algunas plataformas, por ejemplo Reddit, requieren un mínimo de calidad del contenido para que se adecue al objetivo de cada sección (*subreddit*), sus reglas internas y la funcionalidad específica de dicho sitio.[110] Otras plataformas como Wikipedia requieren que las publicaciones cumplan con requisitos de neutralidad, calidad en las fuentes, evitar el plagio y que sean enciclopedicamente relevantes.[111]

El rol de las reglas de moderación de contenidos es muy importante, porque funciona como la fuente que permite evaluar lo que las plataformas se comprometen a hacer (son las reglas que están obligadas a seguir), por un lado, y porque además permiten evaluar si ellas se corresponden o no con ciertos estándares mínimos, como los de transparencia y los del derecho a la libertad de expresión.

[108] Gillespie, Tarleton. "Custodians of the...", *op. cit.*. p. 46.
[109] Idem.
[110] Ibid, p. 64.
[111] Wikipedia. "Políticas y convenciones". Consultado el 15 de octubre de 2021 en: https://es.wikipedia.org/wiki/Wikipedia:Pol%C3%ADticas_y_convenciones.

c. Procedimientos de Moderación

Existen distintos momentos en los que se realiza la moderación de contenidos:

> **1. Moderación *ex ante:*** los moderadores humanos o los sistemas automatizados revisan el contenido antes de que sea publicado.[112]

Este método también es conocido por los filtros de subida que tienen algunas plataformas como Twitter o Facebook para revisar principalmente que lo que se va a publicar no es pornografía infantil o implica el uso "indebido" de algún material protegido por derechos de autor.

> **2. Moderación *ex post*:** la revisión de los contenidos se hace de manera posterior a que han sido difundidos

Después de la publicación del contenido, las expresiones pueden ser revisadas por los moderadores en caso de que otra persona usuaria haga *flagging*,[113] o que un moderador humano, sistema automatizado o terceros realicen una denuncia de dicho contenido.[114] Algunas plataformas como Facebook poseen un sistema diseñado para filtrar denuncias realizadas por parte de los usuarios.

La moderación posterior puede realizarse de distintas formas:

- **Moderación Reactiva:** el moderador revisa contenido de manera pasiva o hace moderación solo hasta que un tercero se lo pide.

  La moderación reactiva es el método más usado por muchas plataformas en cuanto a su moderación de contenido.[115] Algunas plataformas usan sistemas de flagging para que sean los mismos usuarios quienes de manera comunitaria denuncien un post que consideran inadecuado.

[112] Klonick, Kate. "The New Governors...", *op. cit.*, p. 91.
[113] Mecanismo por el cual los mismos usuarios de las plataformas pueden denunciar contenido que consideren inapropiado para que sea revisado por algún moderador.
[114] Roberts, Sarah. "Behind The Screen..". *op.cit.*, p. 33.
[115] Klonick, Kate. "The New Governors..." *op. cit.*, p. 1638.



La mayoría de las denuncias se filtran en un primer momento para evitar la carga de trabajo innecesaria para las distintas etapas de la moderación. Por ejemplo, en Twitter o Facebook piden que la persona usuaria categorice su denuncia para ver si de entrada sería procedente o si solo fue del desagrado personal del usuario.

- **Moderación Activa:** los moderadores buscan proactivamente contenido que no cumpla con los términos y condiciones.[116]

La moderación activa es el método que se usa normalmente frente a ciertos discursos que no son protegidos por el derecho a la libertad de expresión, y puede ser tanto automatizada como manual. Un ejemplo de dicha moderación es la que se hace frente a expresiones de grupos terroristas o extremistas.

Tanto la moderación reactiva como la activa pueden ser realizadas por dos tipos de moderadores:

- **Humanos:** personas moderadoras a las que se les da un entrenamiento para revisar distintos contenidos dependiendo de la plataforma en la que se encuentren.
- ***"Algoritmos"*:** sistemas automatizados que usan aprendizaje de máquina (conocido como *machine learning*) al que se le entrenó para buscar ciertos contenidos y eliminarlos si incumplen con los términos y condiciones.[117]

Es importante señalar que la moderación automatizada posee el dilema de que las expresiones en general se emiten en un contexto determinado que incorpora cuestiones que deben valorarse, tales como la intencionalidad y las circunstancias históricas, culturales y sociales del caso concreto, por lo que muchas veces los contenidos pueden ser suprimidos o moderados de manera incorrecta por los sistemas dada la incapacidad que tienen para tomar en cuenta ese contexto.[118]

[116] Id. Los errores que surgen de la descontextualización y de la interpretación de las distintas variables a tomar en cuenta (y también de los sesgos inherentes a los sistemas) pueden potenciar sin duda la discriminación contra grupos vulnerables. Al respecto ver:Del Campo, Agustina; Schatzky, Morena; Hernández, Laura; Lara, Juan Carlos. *Mirando al Sur. Hacia nuevos consensos regionales en materia de responsabilidad de intermediarios en Internet*, Al Sur, Abril 2021, p. 31.
[117] Idem.
[118] Roberts, Sarah. "Behind The Screen..", *op.cit.*, p. 34.

El dilema contextual de las expresiones es lo que explica la necesidad de que existan mecanismos de apelación y de revisión de la moderación llevada a cabo por los sistemas automatizados, y de cuestionar que la moderación de contenidos sea llevada a cabo por ellos únicamente.

d. Efectos de la moderación en la libertad de expresión y otros derechos

El principal dilema del debate de la moderación de contenidos es la dificultad para llegar a un equilibrio óptimo entre una moderación casi inexistente o una moderación que sea desproporcionada y termine lesionando intereses fundamentales como el de la libertad de expresión.

A continuación hablaremos de los dos extremos del péndulo y sus consecuencias, principalmente aquellas que afectan en la libertad de expresión pero también en derechos como la no discriminación, el derecho a la integridad física y mental, así como los derechos de las niñas, niños y adolescentes.

i. Consideraciones prácticas y límites de los distintos modelos de moderación: ¿Qué pasa si no se modera?

Es imposible señalar que existe una plataforma o sitio en internet que no realiza algún tipo de moderación de contenidos. En particular, porque existen mandatos expresos en ley para eliminar contenidos que no están protegidos por la libertad de expresión, tal como señalamos anteriormente.

La moderación existe y no debe discutirse en torno a ella como si se tratara de un problema sencillo. Las principales consecuencias de no moderar pueden englobarse de la siguiente manera:

- **Spam**
  Son aquellos mensajes automatizados o coordinados no solicitados, que se envían repetidamente con el fin de acaparar la atención de un usuario inundando los canales informativos.[119]

[119] Internet Society. *What Is Spam*. Consultado el 31 de diciembre de 2018. Disponible en: https://www.internetsociety.org/resources/doc/2014/what-is-spam/#_ftn2.



El contenido del spam puede ser desde una estrategia para vender un producto hasta una estrategia criminal para acceder a los datos personales o cuentas de las y los usuarios.

- **Acoso, amenazas y violencia en línea**

  La violencia en línea es un fenómeno constante en las plataformas, en especial contra grupos en situación de vulnerabilidad. Un dato que refleja esta situación es que las mujeres jóvenes de entre 18 y 30 años son las más afectadas por estas agresiones, así como el hecho de que el 40% de estas violencias son cometidas por personas conocidas por las sobrevivientes y 30% por desconocidos.[120]

  La dinámica de grupos exacerba este tipo de conductas,[121] y puede generar ataques coordinados a sus víctimas, mismos que inevitablemente obstaculizan o vulneran el derecho a la libertad de expresión de las personas o grupos sujetos de acoso.[122]

- **Pornografía infantil**

  Como se mencionó en el apartado de límites a la libertad de expresión, la pornografía infantil no está protegida por este derecho, sino que se trata de un discurso prohibido de manera expresa.

  Existen numerosos casos que muestran el problema real de la pornografía infantil y las medidas que tanto autoridades como intermediarios han tomado para combatirla. En Estados Unidos, por ejemplo, se vió un incremento en casos de explotación infantil y pornografía infantil desde el año 2012.[123] Tanto el FBI y el Departamento de Justicia, como distintas plataformas han colaborado para generar estrategias para contrarrestar estos delitos.

[120] Luchadoras, Artículo 19, Cimac, et. al. “La violencia en línea en contra las mujeres en México”, 2017, p. 16. Disponible en: https://r3d.mx/wp-content/uploads/180125-informe_violencia_en_linea_mx-v_lanzam.pdf.

[121] Sunstein, Cass. *Republic.com 2.0*, Princeton University Press, 2009. p. 60.

[122] Keats Citron, Danielle. *Hate Crimes in Cyberspace*, Harvard University Press, United States of America, 2014. pp. 193-197.

[123] Dube Ryan. “Unfortunate Truths about Child Pornography and the Internet”, Make Use Of. Consultado el 7 de diciembre de 2012. Disponible en: https://www.makeuseof.com/tag/unfortunate-truths-about-child-pornography-and-the-internet-feature/.

Cabe recalcar que la legislación estadounidense hace una excepción al principio de inmunidad de intermediarios en cuestiones relacionadas con la pornografía infantil. Por ello se han implementado sistemas automatizados para la detección de imágenes de pornografía infantil.

Un ejemplo de estos sistemas es el *Photo DNA* elaborado por Microsoft.[124] *Photo DNA* usa una huella digital (*hashing*) de imágenes que pueden ser correlacionadas con una base de datos con fotografías recopiladas de bases de pornografía infantil. Esto facilita la detección de páginas o publicaciones que contienen dichas imágenes y ayuda a la persecución de este delito.

- **Contenido sexual explícito**

  Las discusiones en torno a la publicación de contenido sexual explícito son centrales en la moderación de contenidos. Sitios como Youtube o Instagram han decidido que su plataforma no albergue contenidos pornográficos. Por ejemplo, Youtube prohíbe cualquier contenido sexual explícito que genere gratificación sexual, pero permite contenido con desnudos cuando sea con fines educativos, artísticos o de salud (de manera similar a como hace Instagram), y puede restringir el contenido que no sea sexualmente explícito pero tenga "insinuaciones sexuales" para ciertas edades.[125]

  Existen áreas grises, tal como lo señala Youtube al tratar de definir contenido sexual, así como también ciertas problemáticas que pueden sobre-censurar contenido protegido por la libertad de expresión (las analizamos en el siguiente apartado).

  La lógica detrás de este tipo de medidas responde a que muchas plataformas quieren acceder a ciertos grupos de personas en específico, y a que la inclusión de contenido sexual explícito podría alejar a muchas personas usuarias de su plataforma.

[124] International Centre For Missing and Exploited Children. "Giving law enforcement the tools it needs to fight child sexual exploitation". Disponible en: https://www.icmec.org/train/law-enforcement/technology-tools/.
[125] Youtube. "Nudity and Sexual Content Policies". Disponible en: https://support.google.com/youtube/answer/2802002?hl=en&ref_topic=9282679#zippy=%2Cother-types-of-content-that-violate-this-policy%2Cage-restricted-content.



La difusión de contenido sexual sin consentimiento también es un tema preocupante que las plataformas deben de atender. Algunas plataformas como Twitter o Facebook tienen un medio de apelación para cuando estos contenidos son subidos, sin embargo hace falta que dichos mecanismos sean expeditos, transparentes y efectivos.[126]

- **Violencia gráfica o explícita**

  Hay un consenso bastante generalizado en las plataformas de redes sociales sobre la prohibición de contenidos violentos explícitos. Si bien el grado varía de plataforma a plataforma, en general el objetivo es evitar que estos contenidos se utilicen para fomentar la violencia.[127]

  Por ejemplo, organizaciones terroristas o del crimen organizado han usado el recurso de difusión en redes sociales como propaganda para reclutar miembros, así como para dar a conocer su letalidad, fortaleza y recordarle al público su poder.[128] Un caso en donde el crimen organizado tiene una presencia importante es el de México, en donde esto sucede con los cárteles de narcotráfico. Por ejemplo, recientemente se documentó la estrategia de comunicación que varios grupos del crimen organizado mantienen en la plataforma de *Tik Tok*.[129]

  No obstante, este criterio tampoco existe sin matices. En el siguiente apartado analizaremos la dificultad de realizar la revisión de publicaciones bajo esos criterios amplios. En especial, cuando persiguen un objetivo legítimo al atender al interés público o proteger otros derechos humanos.

[126] Para más información ver: "Keats Citron Danielle "Hate crimes in cyberspace", *op. cit.*; Goldberg, Carry. "Nobody's Victim", op. cit.; Luchadoras MX, Artículo 19, APC, et. al. "La Violencia en línea contra las mujeres en México". Disponible en: https://r3d.mx/wp-content/uploads/180125-informe_violencia_en_linea_mx-v_lanzam.pdf.

[127] Gillespie, Tarleton. "Custodians of the...", *op. cit.*, pp. 54-55.

[128] Fernandez M., Alberto. "Here to stay and growing: Combating ISIS Propaganda Networks", Brookings Institution. Disponible en: https://www.brookings.edu/wp-content/uploads/2016/07/IS-Propaganda_Web_English_v2-1.pdf.

[129] Lopez, Oscar. "Los cárteles mexicanos invaden Tik Tok", The New York Times. Consultado el 28 de noviembre de 2020. Disponible en: https://www.nytimes.com/es/2020/11/28/espanol/america-latina/cartel-tiktok.htm.

ii. Los efectos de la vaguedad o ambigüedad de los criterios de moderación

No existe ninguna moderación que no tenga que lidiar con matices, y previamente señalamos cómo algunos criterios como "contenido sexual explícito" o "contenido violento" son conceptos cuya interpretación puede ser controvertida y que generalmente recae en la plataforma la resolución de dicha controversia.

Por esta razón es particularmente importante discutir sobre los matices y, cuando los haya, los casos en los que la moderación debe tomar en cuenta elementos particulares que eviten llegar a una moderación que termine afectando la libertad de expresión de las personas usuarias de las plataformas. Revisemos algunas de estas situaciones.

- **Contenido gráfico o violento de interés público**

    En este caso, la excepción a la regla de la prohibición de difusión de contenidos de violencia se manifiesta claramente cuando un video es de interés público. Si bien es contenido impactante o que puede resultar desagradable para muchas personas, su relevancia se encuentra -por ejemplo- en denunciar crímenes de lesa humanidad (o de otro tipo) que sean acallados por gobiernos u otros sujetos.

    Un ejemplo de este tipo fue el primer video de la guerra de Siria, subido a Youtube en el año 2011, y que consistía en un video que muestra el cuerpo del adolescente Hamza al-Khatib golpeado y quemado. Hamza había sido detenido mientras acudía a la protesta en contra del gobierno de Bashar al-Assad, por lo que este video desató la indignación pública y el adolescente fue el símbolo de la Revolución Siria. Sin embargo, Youtube decidió removerlo de su plataforma por ir en contra de la política de contenido gráfico.

    Esta decisión fue cuestionada por el alto interés público que suponían dichas imágenes para las personas en Siria, poniendo en duda qué clase de poder tenían estas plataformas para decidir cuando algo era relevante para la sociedad y cuando no. A raíz de las múltiples protes-



tas contra la decisión de la plataforma, Youtube decidió mantener el video dentro de su plataforma con un filtro de restricción de edad.[130]

Otros ejemplos son visibles en el caso de México, donde también se han reportado controversias parecidas, tal como la del video publicado en Facebook donde se mostraba la ejecución a una maestra hecha por un cártel del narcotráfico.[131]

Este video fue compartido por distintos usuarios para condenar la violencia.[132] El interés público en ellos estaría en refutar la postura negacionista del gobierno mexicano ante la violencia del narcotráfico. Los otros videos de denuncias fueron subidos por la ciudadanía o periodistas que usaban las redes para denunciar la violencia que se seguía viviendo en el norte del país. Dicha violencia fue desestimada por la Presidencia de la República alegando que la ciudadanía sufría de "histeria colectiva". Por lo tanto, la única vía disponible para denunciar la situación de violencia fue a través de redes sociales.

Sin embargo, en otros países el video de la ejecución de la maestra fue criticado porque era innecesariamente gráfico para una plataforma como Facebook donde podían verlo personas altamente impresionables. Si bien en un primer momento Facebook había decidido mantenerlo en la plataforma, posteriormente resolvió removerlo de la misma.[133]

- **Contenido sexual o desnudos**

Definir qué es contenido sexual explícito es complicado, por decir lo menos. Si bien puede entenderse que una plataforma no busque albergar contenido pornográfico para poder llegar a cierto público en específico, es pertinente preguntar, ¿cuáles son los estándares para considerar contenido

[130] Kaye, David. "Speech Police. The Global Struggle to Govern the Internet", Columbia Global Reports, 2019, pp. 22-23.
[131] Grant, Will. "Facebook beheading video: Who was Mexico's Jane Doe?", BBC News Consultado el 4 de Noviembre de 2013. Disponible en: https://www.bbc.com/news/magazine-24772724.
[132] Kelion, Leon. "Facebook lets beheading clips return to the social network", BBC News, Consultado el 23 de Octubre de 2013. Disponible en: https://www.bbc.com/news/technology-24608499.
[133] Memott, Mark. "Facebook removes beheading video, says it will tighten rules". NPR. Consultado el 23 de octubre de 2013. Disponible en: https://www.npr.org/sections/the-two-way/2013/10/23/240190936/facebook-removes-beheading-video-says-it-will-tighten-rules.

sexual explícito o no? ¿Qué pasa, por ejemplo, con las trabajadoras sexuales que ofrecen servicios en algunas plataformas porque es una vía más segura que hacerlo en la calle?

En cuanto a la primera pregunta sobre qué es contenido sexual explícito o no, algunas plataformas hacen un catálogo con partes del cuerpo que son exhibidas, otras señalan que se trata del contenido que busca generar satisfacción sexual en las personas y algunas realizan catálogos de partes del cuerpo que consideran "inapropiadas" de exhibirse en su plataforma.

De lo anterior, es inevitable mencionar el elevado nivel de subjetividad que existe al momento de moderar este tipo de contenidos. Una persona que tenga criterios o percepciones más conservadoras puede censurar expresiones protegidas por la libertad de expresión, el libre desarrollo de la sexualidad o incluso el derecho a la salud.

Un ejemplo de los problemas de este nivel de subjetividad es el de los grupos de activistas que promueven la normalización del amamantamiento. En el año 2008, subieron fotos dando el pecho a sus bebés en sus perfiles en Facebook como parte de una protesta virtual en contra de la política que prohibía la publicación de imágenes de amamantamiento que tuvieran pezones visibles.[134]

En el año 2015,[135] Facebook volvió a remover la imagen de una madre amamantando a su hijo, situación que derivó nuevamente en reclamos de las usuarias y terminó con la plataforma modificando sus normas comunitarias para explicitar que las imágenes de amamantamiento están permitidas en sus plataformas,[136] esta vez sin importar qué parte del pecho se exhiba en las publicaciones.

- **Discurso de odio, discursos polarizantes o chocantes**

  Los discursos estigmatizantes hacia grupos vulnerables son reprochables y pueden terminar afectando derechos de las personas que in-

[134] Sweeney, Mark. "Mums furious as Facebook removes breastfeeding photos", The Guardian. Consultado el 30 de diciembre de 2008. Disponible en: https://www.theguardian.com/media/2008/dec/30/facebook-breastfeeding-ban.
[135] Idem.
[136] Facebook. "¿Permite Facebook la publicación de madres amamantando?". Consultado el 20 de febrero de 2021. Disponible en: https://www.facebook.com/help/340974655932193.



tegran dichos grupos. Sin embargo, estos discursos no se pueden limitar a *palabras prohibidas*, ya que el lenguaje evoluciona de manera contínua y se utiliza también de distintas formas según el contexto; es decir, que las palabras carecen de significado cuando son sacadas de su contexto en específico.

Por ejemplo, existen palabras raciales despectivas que tienen un homónimo a otro concepto cotidiano. También hay palabras insultantes hacia grupos vulnerables que han sido apropiadas por los mismos, tales como los insultos hacia la comunidad LGBTTTIQ+ que ahora son utilizadas entre las personas que integran estos grupos.

Otro ejemplo es que existen ocasiones donde se pueden usar palabras controversiales dentro de espacios artísticos, para fines periodísticos o incluso en ocasiones donde el insulto ya está despegado de su concepto original, por lo que no producen efectos estigmatizantes sobre grupos vulnerables.

iii. La relación entre la concentración y el impacto en los derechos humanos

El impacto de las decisiones de moderación de contenidos en línea en el derecho a difundir, recibir o buscar información depende, en gran medida, de la existencia de alternativas para que las personas usuarias puedan ejercer su derecho a la libertad de expresión en una plataforma o servicio distinto.

Es decir, difícilmente podría argumentarse que la remoción de una publicación de un intermediario afecta de manera considerable su derecho a la libertad de expresión o el libre flujo informativo si una persona usuaria posee alternativas de publicación con igual o mayor posibilidad de alcance.

Por el contrario, cuando una plataforma dominante en Internet toma una decisión de moderación en su plataforma, la imposibilidad o dificultad de difundir, recibir o buscar determinada información con el mismo alcance a través de otra plataforma incide de manera determinante con el derecho a la libertad de expresión.

Es por ello que al analizar la realidad compleja de la moderación de contenidos en Internet, es fundamental diferenciar entre la moderación de contenidos llevada a cabo por plataformas que por su tamaño u otra circunstancia pueden limitar de manera trascendental el alcance de una expresión, y la moderación de contenidos llevada a cabo por intermediarios sin ese poder.

e. La dificultad de la moderación a escala

Hasta ahora se ha descrito la problemática general que resulta de la moderación de contenidos en las plataformas, sin embargo, no todas las plataformas tienen el mismo tipo de moderación y más importante: no todas las plataformas son grandes plataformas.

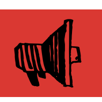
Cuando el poder legislativo sugiere combatir el contenido "malicioso" que se encuentra en las plataformas, muchas veces se refieren únicamente a las plataformas dominantes en el mercado como Facebook o Google. Lamentablemente, la miopía de reducir a Internet a unas cuantas plataformas puede provocar o agravar barreras a la competencia para plataformas emergentes, en favor de empresas dominantes que en sus inicios se beneficiaron de la ausencia de este tipo de regulación estricta, lo que les permitió crecer y obtener la posición dominante de la que gozan actualmente.

Para evitar los obstáculos que limitan la competitividad y afectan a los usuarios de plataformas digitales, es necesario entender los principales problemas de la moderación a escala y entender que la moderación es un juego de suma cero: siempre habrá alguien que termine inconforme con la decisión final del moderador.[137]

El Teorema de imposibilidad de Masnick[138] establece que la moderación de contenidos a gran escala es imposible de hacer a la perfección (al 100%). Masnick

[137] Goldman Eric, Miers Jess. "Why Internet Companies Can't Stop Awful Content", Social Science Research Network. Rochester, NY. 1 de enero de 2020. p. 3. Disponible en: https://doi.org/10.2139/ssrn.3518970.

[138] Masnick Mike. "Masnick's Impossibility Theorem: Content Moderation At Scale Is Impossible To Do Well", Techdirt. Consultado el 5 de octubre de 2021. Disponible en: https://www.techdirt.com/articles/20191111/23032743367/masnicks-impossibility-theorem-content-moderation-scale-is-impossible-to-do-well.shtml.



argumenta que en la moderación siempre habrá alguien que gana y alguien que pierde por lo que nunca habrá un escenario en el que todas las personas estén satisfechas con el resultado. El dilema se vuelve más complicado a medida de que va incrementando el número de usuarios y publicaciones que moderar:

> *Conseguir el 99.9% de las decisiones en un nivel de resultados aceptable quizá funcione en situaciones donde no estás trabajando con más de 1,000 decisiones de moderación al día, pero las grandes plataformas tienen que lidiar con mucho más que eso" [...] "Si asumes que hay un millón de decisiones correctas al día, incluso con el 99.9% de efectividad, la plataforma sigue teniendo 1,000 decisiones incorrectas.*[139]

Por otro lado, la infraestructura técnica y humana para realizar la moderación de contenidos es costosa. 350 millones de fotos son subidas a diario en Facebook, sin contar otro tipo de publicaciones.[140] Es necesario un ejército de moderadores y sistemas automatizados que soporten estas cantidades masivas de información cada minuto para que la plataforma pueda funcionar.

Por ejemplo, Facebook tiene a 30,000 personas trabajando en seguridad en la plataforma, de las cuales 15,000 son moderadores con contratos de tiempo completo, sin contar a las personas que contrata como prestadoras de servicio para que pueda "estar a la escala global". El pago promedio de un empleado en Facebook es de 240,000 dólares al año, mientras que el de un prestador de servicios es solamente 28,800 dólares al año. Facebook es una empresa que en el 2019 reportó que ganaba 6.9 billones de dólares al año en ingresos.[141] Incluso Zuckerberg anunció en 2019 que invertiría más de 3.7 billones de dólares para temas de seguridad en la plataforma, e incluso mencionó que era mucho más que el total de las ganancias anuales de Twitter.[142]

[139] Idem.
[140] Cooper Smith. "Facebook Users Are Uploading 350 Million New Photos Each Day", Business Insider. Consultado el 12 de octubre de 2021. Disponible en: https://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9.
[141] Newton, Casey. "The Secret Lives of Facebook Moderators in America", The Verge. Consultado el 25 de febrero de 2019. Disponible en: https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona.
[142] Lauren Feiner Rodriguez, Salvador. "Mark Zuckerberg: Facebook Spends More on Safety than Twitter's Whole Revenue for the Year", CNBC. Consultado el 23 de mayo de 2019. Disponible en: https://www.cnbc.com/2019/05/23/facebook-fake-account-takedowns-doubled-q4-2018-vs-q1-2019.html.

Precisamente, que Facebook pueda alardear sobre su presupuesto de seguridad contra las ganancias totales de otra plataforma es evidencia de las disparidades que existen incluso entre las plataformas más grandes. Por lo cual, no se deben de establecer criterios de moderación que solo sean feasibles para empresas tan grandes como Facebook o Google. Las grandes plataformas pueden pagar los costos para cumplir con sus nuevas obligaciones legales porque en primer lugar fueron hechas pensando en esas plataformas, pero estas reglas supondrían una barrera de entrada al mercado para futuras plataformas emergentes o proyectos alternativos de espacios sociales que se centren en contenido generado por personas usuarias.[143]

Otro problema de las propuestas de regulación que consideran a las plataformas como homogéneas es que olvidan que la moderación es subjetiva y existen distintos tipos de moderación. Es un error común pensar que todos los procesos de moderación se realizan de manera automática o por el ejército de moderadores que están revisando cada una de las publicaciones cada segundo. Existen plataformas como Reddit, donde tienen lineamientos generales para toda su plataforma y tienen equipos para revisar las faltas de un subgrupo,[144] pero también cuentan con moderadores voluntarios o voluntarias en cada subreddit que se dedican a vigilar que las personas afiliadas a ese subgrupo cumplan con las reglas acordadas de ese espacio en específico.[145]

Las principales razones para considerar una regulación diferenciada con base en el tamaño pueden ser para (1) responsabilizar a las empresas grandes que han realizado el mayor daño, (2) reducir la barrera de entrada a competidores y (3) por buscar la equidad en un mercado de competidores con grandes diferencias de ganancias y tamaño.[146]

La actual propuesta de la *Digital Services Act* (DSA) que se discute en Europa, sí contempla una excepción a las micros y pequeñas empresas intermedia-

[143] Eric Goldman, Jess Miers. "Why Internet Companies..." *op. cit.*, p. 4.
[144] Reddit. "Content Policy - Reddit". Consultado el 5 de octubre de 2021. Disponible en: https://www.redditinc.com/policies/content-policy.
[145] También estos moderadores voluntarios tienen lineamientos a seguir que son impuestos por la plataforma. Al respecto ver: https://www.redditinc.com/policies/moderator-guidelines.
[146] Goldman, Eric; Miers, Jess. "Regulating Internet Services by Size", SSRN Scholarly Paper, Social Science Research Network. Rochester, NY. Consultado el 1 de mayo de 2021, p. 2. Disponible en: https://papers.ssrn.com/abstract=3863015.



rias de cumplir con lo establecido en la ley. La regulación tiene el fin de evitar cargas desproporcionadas sobre las empresas emergentes salvo que estas empresas tengan el alcance o impacto parecido a una plataforma grande.

La DSA establece responsabilidades mayores sobre las plataformas muy grandes, y define como "plataforma muy grande"[147] a aquellas que tienen un número mensual de usuarios activos en la Unión Europea igual o mayor a 45 millones. Esto lo hace bajo el principio de proporcionalidad, por lo que si bien los requerimientos son mayores y más estrictos también es cierto que las grandes empresas cuentan con el presupuesto y la infraestructura para realizar el cumplimiento de estas obligaciones.[148]

Existen distintos tipos de métrica para el tamaño de una plataforma:[149]

- Por edad de la empresa
- Por número de empleados
- Por su capitalización del mercado
- Por sus ingresos
- Por el consumo de usuarios: que a su vez se puede dividir en el consumo de usuarios al mes, en los usuarios registrados o en la visita de las páginas

Goldman y Miers consideran que no hay una respuesta categórica para decir qué métrica en específico debe usarse, ya que cada una puede tener sus desventajas si es aplicada de manera categórica. Sin embargo, los autores proponen que se debe de tomar en cuenta los siguientes factores:[150]

1. La métrica debe ser publicada y auditada constantemente.
2. La métrica debe tener una clara definición de la organización, límites materiales y límites económicos que constituyen cada plataforma. Por

[147] European Commission, "Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC" (2020), artículo 25, párr. 1, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services.

[148] European Commission, "Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC" (2020), 7–11, https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-parliament-and-council-single-market-digital-services-digital-services.

[149] Goldman, Eric; Miers, Jess. "Regulating Internet Services by Size.." *op.cit.* pp. 2-3.

[150] Id., pp. 4-5.

ejemplo, saber sobre qué estructura corporativa se está midiendo (ejemplo: no es lo mismo medir solamente el servicio de Google como Google, que buscar medirlo junto con sus otros servicios como Gmail, Alphabet, etc.). Así como no se puede medir solo por el contenido generado por usuarios sin considerar la capitalización del mercado de la plataforma. El ejemplo claro es Wikimedia, que puede ser considerado como una plataforma enorme por la gran cantidad de contenido usado por sus editores y demás personas usuarias pero no recibe tantos ingresos y tiene un equipo muy reducido.

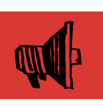
3. El periodo de la medida. Los reguladores deben especificar el margen temporal sobre el que se va a realizar la métrica.
4. Usar distintas métricas para evitar la volatilidad del mercado y los falsos positivos.

f. Aspectos jurisdiccionales de la moderación de contenidos

En este trabajo hemos referido a un marco general de estándares que tienen una aplicación diversa en las jurisdicciones locales y que pueden cambiar según la rigidez de cada sistema jurídico. Además, algunos gobiernos buscan influenciar o regular las plataformas para pedir la remoción de contenidos no solo de su jurisdicción sino de todos los demás países, apoyados en su legislación, en procesos judiciales o al influenciar a las plataformas con su poder público. Por ejemplo, en noviembre del año 2006, el gobierno tailandés anunció que bloquería Youtube a todas las personas usando una IP tailandesa a menos que Google removiera 20 videos que iban en contra de una ley que establecía la prohibición de insultar al rey (con pena de 15 años de prisión).[151]

Para Nicole Wong, trabajadora de Google, fue un shock cultural porque mientras algunas imágenes eran claramente en contra de las normas comunitarias, había otras caricaturas que solamente eran imágenes manipuladas en Photoshop. No obstante, en el contexto cultural de Tailandia existe un amplio cariño por el rey, por lo que decidieron remover los videos dentro de las fronteras geográficas de Tailandia.[152]

[151] Klonick, Kate. "The New Governors...", *op. cit.*, p. 1623.
[152] Idem.



En Turquía sucedió otro incidente similar: un programa de parodia insinuó que Mustafa Kemal, el fundador de la Turquía moderna, era homosexual. Por lo cual, un juez ordenó que se bloqueara el acceso a Youtube a todas las personas usuarias turcas. Si bien el video después fue removido de manera voluntaria, el gobierno le exigió a Google que removiera de la plataforma varios videos ofensivos más.[153]

Google resolvió remover los videos que la compañía consideraba que efectivamente violaron la ley turca, pero únicamente de la jurisdicción turca. Un año después, el gobierno de Turquía le exigió que prohibiera el acceso a estos videos en todo el mundo, Google se negó y el gobierno de Turquía bloqueó el acceso a la plataforma de Youtube en todo el país.[154]

Evidentemente existen problemas logísticos para las plataformas al tratar de implementar distintos modelos de moderación en diversos países. Muchas de las plataformas preponderantes que se desarrollaron en EUA tienen como base los principios de la Primera Enmienda, pero cuando estas empresas alcanzaron niveles globales, descubrieron que los filtros de contenido para cada jurisdicción en específico complican la moderación, por lo que decidieron programar un mismo conjunto de estándares y expandirlos según vayan ocurriendo de acuerdo a las presiones legales del gobierno en turno.[155]

Cuando una plataforma decide imponer categóricamente una serie de principios de moderación globalmente, sin tomar en cuenta el contexto de derechos humanos y cultural de cada país, provoca una serie de afectaciones a los derechos de la libertad de expresión y del acceso a la información.

El caso de *Google Inc. vs Equusteck solutions Inc.* de la Suprema Corte de Canadá sirve para ejemplificar cómo la decisión de un país afecta el derecho a la libertad de expresión. En este caso, la disputa se originó debido a que la empresa Equustek demandó a la empresa Datalink por adjudicarse uno de sus productos, por lo cual había una violación a la propiedad intelectual de la primera empresa. Equustek le requirió a Google que desindexara las páginas de Datalink que se usaban para hacer negocios en línea. Después de que un

[153] Ibid, p. 124.
[154] Idem, citando a Jeffrey Rose., "The Delete Squad", New Republic, 29 de Abril 29 de 2013.
[155] Keller, Daphne. "Who do you Sue…". *op. cit.*, p. 8.

tribunal ordenó a Datalink dejar de operar y hacer negocios en línea, Google removió los links de sus dominios canadienses pero se rehusó a remover los resultados del dominio global. Equustek solicitó a través de un tribunal una orden interlocutoria para que se desindexara a nivel global; Google apeló ante la Suprema Corte de Canadá, misma que falló en favor de la empresa.[156] Los razonamientos de la Corte candiense fueron los siguientes:

1. Google debía cumplir con lo ordenado para que dejara de facilitar el daño que provocó Datalink a Equustek. La corte canadiense argumentó que una corte puede ordenar una medida que sea vinculante para la conducta del infractor en cualquier lugar del mundo porque "el internet no tiene fronteras, su hábitat natural es global".[157] Por lo cual, la corte canadiense decidió que el mandato judicial debía tener un impacto global para asegurar su efectividad.

2. Si bien Google argumentó que la decisión de remover el contenido internacionalmente podría resultar en responsabilidad internacional por parte del Estado canadiense al violar la jurisdicción de otros estados y afectar la libertad de expresión, la Corte desechó este argumento considerando que era un supuesto teórico pues "la mayoría de los países reconocerían la violación de los derechos de propiedad y verían la responsabilidad legal en vender productos pirata".[158]

3. Respecto al tema de la violación a la libertad de expresión, la Corte decidió que en el caso de que Google tuviera evidencia de que dicha orden judicial violaba las leyes de otra jurisdicción, incluyendo el derecho a la libertad de expresión, podía consultar con las cortes de la Columbia Británica para mdificar la orden al caso concreto.[159]

La decisión fue ampliamente criticada por resultar en una clara intervención a la jurisdicción de otros países y en la afectación a derechos

[156] "Google Inc v. Equustek Solutions Inc. (Equusteck I)", Global Freedom of Expression. Consultado el 11 de octubre de 2021. Disponible en: https://globalfreedomofexpression.columbia.edu/cases/equustek-solutions-inc-v-jack-2/.
[157] Google Inc. v. Equustek Solutions Inc., 2017 SCC 34, párr. 41 (Suprema Corte de Canadá, decidido el 28 de junio de 2017).
[158] Ibid, párr 44. Traducción propia.
[159] Ibid, párras. 45-48.



humanos como la libertad de expresión y acceso a la información.[160] En definitiva, la corte canadiense creó un precedente donde se favorece el interés comercial sobre la libertad de expresión en distintas jurisdicciones y puede utilizarse para justificar restricciones a derechos humanos a escala global.

Por estas razones, la decisión fue combatida en la Corte del Distrito de California del Norte, donde se resolvió que el mandato ordenado por la Corte Canadiense no podía ser aplicado en EUA, por la inmunidad que la Sección 230 de la CDA otorga a Google en territorio estadounidense.

Las plataformas, por sí mismas, también llevan a cabo remociones globales que afectan a la libertad de expresión. Un ejemplo estadounidense es el caso *Yahoo! Inc. v La Ligue Contre Le Racisme et L'Antisémitisme*, o el de Sikhs for justice v Facebook. En el caso Yahoo Inc. v La Ligue Contre Le Racisme es de los primeros que articulan este malentendido como "protección de los derechos de las personas usuarias donde se previene que plataformas remuevan discurso estadounidense basándose en leyes extranjeras". El caso discute si una Corte estadounidense puede aplicar una orden de Francia para que el buscador Yahoo deje de mostrar elementos relacionados con el nazismo, no si Yahoo puede o no cumplir con esto voluntariamente.

Keller señala que, de hecho, Yahoo decidió voluntariamente cumplir con la orden del gobierno francés mientras se encontraba litigando el caso ante los tribunales estadounidenses. Por lo cual, que "una compañía que teme que sus activos extranjeros se pierdan o sus empleados sean arrestados o que no quiere perder acceso a un mercado extranjero lucrativo, puede encontrar buenas razones para seguir las órdenes judiciales de cortes extranjeras y hacerlo globalmente si es lo que le pide dicha corte".[161]

[160] Aaron Mackey Ranieri Corynne McSherry, and Vera. "Top Canadian Court Permits Worldwide Internet Censorship", Electronic Frontier Foundation. Consultado el 28 de junio de 2017. Disponible en: https://www.eff.org/deeplinks/2017/06/top-canadian-court-permits-worldwide-internet-censorship; "Global Internet Takedown Orders Come to Canada: Supreme Court Upholds International Removal of Google Search Results - Michael Geist". Consultado el 11 de octubre de 2021. Disponible en: https://www.michaelgeist.ca/2017/06/global-internet-takedown-orders-come-canada-supreme-court-upholds-international-removal-google-search-results.

[161] Keller, Daphne. "Who do you Sue...", *op. cit.*, pp. 8-9.

El caso de *Sikhs for justice v. Facebook* es otro ejemplo en el que una plataforma censura discursos válidos para evitar confrontaciones con la jurisdicción de un país. Sikhs for Justice (SFJ) es una organización de derechos humanos que se dedica a la incidencia para la independencia de Punjab, en la India. La organización tenía una página de Facebook que usaba para su activismo, organizar campañas de incidencia y promover el derecho de la autodeterminación de las personas Sikh en Punjab.[162] En mayo de 2015, Facebook bloqueó la página en India por requerimiento del gobierno indio. Sikhs for Justice le pidió a la plataforma que les regresara la cuenta y diera una explicación por el bloque pero la plataforma se negó. La organización demandó a Facebook por daños argumentando que la plataforma era responsable de discriminación de racial.[163]

Sin embargo, la jueza de la Corte de Distrito desechó las demandas de la organización, al señalar que la sección 230 protege las decisiones de moderación, incluyendo la de no publicar el contenido de SFJ, por lo que no podía considerarse como "discriminatorio", sino una decisión que la plataforma tenía derecho de realizar.

Lo mismo pasó en el caso *Zhang contra Baidu*, donde un grupo de activistas promotoras de la democracia en China demandaron a la empresa de búsqueda Baidu por bloquear en Estados Unidos, a petición del gobierno de China, una variedad de discursos políticos a favor de la democracia en China. Un juez federal decidió de manera controversial que la decisión de las plataformas sobre el contenido que permanecía y se removía de sus páginas estaba protegido por la Primera Enmienda, a pesar de que se usará para censurar discursos en otras jurisdicciones.[164]

La Corte de Justicia de la Unión Europea (CJUE) también ha resuelto casos relevantes respecto a las remociones globales y desindexación de contenidos. Por ejemplo, el Caso Glawischnig-Piszczek contra Facebook Ireland resuelto por la Tercera Cámara de la Corte, que trata sobre la desindexación de contenido ilícito y el alcance territorial de esta decisión.[165]

[162] Sikhs For Justice "SFJ", INC. v Facebook, INC. Case No. 15-CV-02442-LHK (Northern District of California District Court, 13 de noviembre de 2015).
[163] Idem.
[164] "Zhang v. Baidu.Com, Inc.", Global Freedom of Expression. Consultado el 5 de octubre de 2021. Disponible en: https://globalfreedomofexpression.columbia.edu/cases/zhang-v-baidu-com-inc/.
[165] "Glawischnig-Piesczek v. Facebook Ireland Limited", Global Freedom of Expression. Consultado el 10 de octubre de 2021. Disponible en: https://globalfreedomofexpression.columbia.edu/cases/glawischnig-piesczek-v-facebook-ireland-limited/.



Los hechos del caso muestran que en el 2016, un usuario de Facebook compartió en su página personal un artículo de una revista digital que hablaba sobre la política austríaca Glawischnig-Piszczek. El usuario después publicó, en conexión con el artículo, un comentario que la demandante consideró como dañino a su reputación y difamatorio.[166]

La política austríaca demandó a Facebook ante la Corte Comercial de Viena por no eliminar el comentario. La Corte Comercial ordenó a Facebook que no permitiera la publicación o distribución de fotografías de la demandante si estaban acompañadas con el texto exacto o con palabras de equivalente significado que el del comentario original. El tribunal de alzada confirmó esta decisión pero limitó su alcance: solamente se podía remover el contenido idéntico al comentario. El caso llegó a la Suprema Corte de Austria que resolvió enviar a la CJUE el asunto para que interpretara la legislación de la Directiva de Comercio Digital relevante al caso.[167]

La Tercera Sala de la CJUE resolvió que la directiva de comercio digital no impide que un estado miembro pueda solicitar a un proveedor de servicios remover o bloquear contenido que ha sido declarado ilícito o contenido que es igual o equivalente a esa información ilegal. Respecto de la aplicabilidad geográfica de esa decisión, la corte indicó que la directiva no se pronuncia sobre alguna limitación territorial, por lo cual cada estado miembro podía determinar el alcance geográfico de la restricción, siempre y cuando estuviera en el marco del derecho internacional relevante.[168]

La CJUE también se pronunció sobre el alcance de la remoción de contenidos en el caso *CNIL vs Google*. La autoridad de protección de datos personales de Francia (CNIL) multó a Google por no desindexar de manera global la información respecto a una persona.[169]

La Gran Cámara de la Corte Internacional de Justicia de la Unión Europea resolvió que las legislaciones europeas no hacían mención sobre el ámbito

[166] Ibid, párr. 12.
[167] Ibid, párr 14-20.
[168] Ibid, párr. 27-53.
[169] "Google LLC v. National Commission on Informatics and Liberty (CNIL)", Global Freedom of Expression. Consultado el 10 de octubre de 2021. Disponible en: https://globalfreedomofexpression.columbia.edu/cases/google-llc-v-national-commission-on-informatics-and-liberty-cnil/.

geográfico de aplicación para las órdenes de desindexación. La Corte concedió que el "Derecho a la desindexación" no está reconocido globalmente. En este sentido, la Corte Europea enfatizó que este derecho no es absoluto y debe de ponderarse con otros derechos fundamentales de acuerdo al principio de proporcionalidad.

La Corte de Justicia estableció que en principio la desindexación debe ser posible en la jurisdicción de todos los Estados miembros, pero dado que las protecciones a la privacidad no son uniformes en la Unión Europea, era obligación de las cortes nacionales decidir el alcance de la misma. Finalmente, la Gran Cámara no se pronunció sobre si Google nunca podría estar obligado a llevar a cabo una desindexación global, dejando a cada corte nacional decidir si esto es apropiado.



D. TRANSPARENCIA Y RENDICIÓN DE CUENTAS

El acceso a la información ha sido reconocido como un derecho necesario para combatir a la corrupción, conocer de violaciones a derechos humanos -por parte de autoridades Estatales y particulares- y garantizar la transparencia como una herramienta fundamental para todas las democracias.

La transparencia es fundamental sobre todo en asuntos de interés público -incluidos aquellos que involucran las restricciones o violaciones a derechos humanos- porque estos gozan de una protección reforzada del derecho de acceso a la información. Esta garantía suele estar asociada a una responsabilidad por parte de los Estados; sin embargo, día con día se vuelve más importante que las empresas -sobretodo aquellas con una posición preponderante en su sector- cuyas acciones impactan en el goce y disfrute de los derechos humanos, realicen informes periódicos de transparencia públicos para todas las personas.

A la par de la evolución del derecho a la información y la transparencia, también se han diseñado una serie de mecanismos para que los Estados cumplan con sus obligaciones de respetar, proteger y cumplir los derechos humanos y en reconocer el papel de las empresas como organismos especializados de la sociedad que desempeñan funciones especializadas y que deben cumplir todas las leyes aplicables y respetar los derechos humanos.[170]

Bajo una interpretación armónica del deber de proteger contra las violaciones de derechos humanos cometidas por terceros -incluidas las empresas- que tienen los Estados a la luz del derecho de acceso a la información, existe una responsabilidad proactiva que recae sobre las empresas para transparentar aquellos actos u omisiones que tienen un impacto en los derechos humanos.

[170] Naciones Unidas. Principios rectores sobre las empresas y los derechos humanos. 2011. Pág 1.

Sobre las obligaciones de acceso a la información que tienen los intermediarios se han pronunciado partes de distintos organismos internacionales, tales como las relatorías de las Naciones Unidas[171] y de la OEA[172] para la libertad de expresión, quienes concretamente han reconocido que:

> *"los actores privados deben establecer e implementar condiciones de servicio que sean transparentes, claras, accesibles y apegadas a las normas y principios internacionales en materia de derechos humanos, incluyendo las condiciones en las que pueden generarse interferencias con el derecho a la libertad de expresión o a la privacidad de los usuarios. En este sentido, las empresas deben buscar que cualquier restricción derivada de la aplicación de los términos de servicio no restrinja de manera ilegítima o desproporcionada el derecho a la libertad de expresión".*[173]

Además, los deberes de los intermediarios en materia de transparencia son una parte fundamental para conocer de potenciales actos de corrupción y violaciones a derechos humanos por parte de las autoridades Estatales por lo que, de acuerdo con la CIDH, los intermediarios deberían:

> *"tener la protección suficiente para hacer públicas las solicitudes realizadas por agencias del Estado, u otros actores legalmente facultados, que interfieran con el derecho a la libertad de expresión*

[171] Naciones Unidas. Asamblea General. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Frank La Rue. A/HRC/17/27. 16 de mayo de 2011. Párr. 48. Disponible para consulta en: http://ap.ohchr.org/documents/dpage_s.aspx?m=8

[172] Relator Especial de las Naciones Unidas para la Libertad de Opinión y Expresión, Representante de la Organización para la Seguridad y Cooperación en Europa para la Libertad de los Medios de Comunicación y Relator Especial de la OEA para la Libertad de Expresión. 21 de diciembre de 2005. Declaración Conjunta Sobre Internet y sobre Medidas Antiterroristas; Relator Especial de las Naciones Unidas para la Protección y Promoción del Derecho a la Libertad de Opinión y de Expresión y Relatora Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos. 21 de diciembre de 2010. Declaración Conjunta sobre Wikileaks. Punto 5; Relator Especial de las Naciones Unidas (ONU) para la Protección y Promoción del Derecho a la Libertad de Opinión y de Expresión y Relatora Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos de la OEA. 21 de junio de 2013. Declaración conjunta sobre programas de vigilancia y su impacto en la libertad de expresión. Punto 11.

[173] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 112.

o la privacidad de los usuarios. Es una buena práctica, en este sentido, que las empresas publiquen de manera regular informes de transparencia en los que revelan cuando menos, el número y el tipo de las solicitudes que pueden aparejar restricciones al derecho a la libertad de expresión y a la privacidad de los usuarios".[174]

a. Los principios de Santa Clara

Los Principios de Santa Clara (PSC) en su segunda iteración proponen una serie de principios orientados a promover transparencia significativa y rendición de cuentas respecto de la moderación de contenidos llevada a cabo por intermediarios en Internet.[175]

Los PSC proponen una serie de principios fundacionales y operativos, así como principios dirigidos a gobiernos y otros actores estatales. Los principios fundacionales son principios generales y transversales que todas las empresas, independientemente del modelo de negocio, la antigüedad y el tamaño, deben tener en cuenta al llevar a cabo la moderación de contenidos, incluyendo:

1. **Derechos humanos y debido proceso:** Las compañías deben asegurarse de considerar el debido proceso y los derechos humanos en todas las etapas del proceso de moderación de contenidos y hacer pública información sobre cómo dichas consideraciones son integradas.
2. **Reglas y políticas comprensibles:** Las compañías deben publicar reglas y políticas claras, precisas y accesibles en relación a las decisiones de moderación de contenidos.

[174] CIDH, Relatoría Especial para la Libertad de Expresión. Libertad de expresión e internet. OEA/Ser.L/V/II. CIDH/RELE/INF.11/13, 31 de diciembre de 2013, párr. 113. Naciones Unidas. Asamblea General. Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión, Frank La Rue. A/HRC/17/27, 16 de mayo de 2011, párr. 46. Disponible en: http://ap.ohchr.org/documents/dpage_s.aspx?m=85; Global Network Initiative. "A Call for Transparency from Governments and Telecommunications Companies"; Global Information Society Watch. "Don't censor censorship: Why transparency is essential to democratic discourse. Como ejemplo, ver también: Google. Informe de transparencia; Twitter. "Transparency Report. Communicate fearlessly to build trust"; Microsoft. "Law Enforcement Requests Report".
[175] Principios de Santa Clara. The Santa Clara Principles On Transparency and Accountability in Content Moderation. Traducción propia. Disponible en: https://santaclaraprinciples.org/.



3. **Competencia cultural:** Las compañías deben asegurarse de que sus reglas y políticas, así como su observancia, tomen en consideración la diversidad de culturas, lenguages y contextos en los que sus plataformas y servicios se encuentran disponibles y son utilizados.

4. **Riesgos del involucramiento estatal:** Las compañías deben reconocer los riesgos derivados del involucramiento del Estado en la definición y observancia de las reglas y políticas de moderación de contenidos de dichas compañías.

5. **Integridad y explicabilidad:** Las compañías deben asegurarse de que los sistemas de moderación de contenidos, tanto automatizados como no automatizados funcionen con fiabilidad y efectividad. Lo anterior incluye la necesidad de perseguir la precisión y no discriminación en los métodos de detección de contenidos, evaluar regularmente los sistemas para únicamente utilizar aquéllos que provean suficiente confianza de su pertinencia y ofrecer transparencia y supervisión independiente.

Por su parte, los principios operacionales establecen expectativas más detalladas para las empresas más grandes o maduras con respecto a etapas y aspectos específicos del proceso de moderación de contenidos.

En contraste con los estándares mínimos establecidos en la primera iteración (Números, Notificación y Apelación), esta segunda ola de principios proporciona una mayor especificidad, con precisión sobre qué información se necesita para garantizar una transparencia y rendición de cuentas significativas.

Los Principios de Santa Clara 2.0 amplían el alcance de la transparencia que se requiere con respecto a lo que se considera como "contenido" y "acción" de una compañía. El término "contenido" se refiere a todo el contenido generado por las personas usuarias, pagado o no, en un servicio, incluida la publicidad. Los términos "acción" y "actuado" se refieren a cualquier forma de observancia del cumplimiento tomada por una compañía con respecto al contenido o la cuenta de una persona usuaria debido al incumplimiento de las reglas y políticas, incluida (entre otras) la eliminación de contenido, la disminución de la visibilidad algorítmica de contenido y la suspensión (temporal o permanente) de cuentas.

Los PSC proponen 3 principios operacionales:

1. Números (Transparencia). Las empresas deben publicar la cantidad de publicaciones eliminadas y las cuentas suspendidas de forma permanente o temporal debido a violaciones de sus políticas de contenido.

2. Notificación. Las empresas deben notificar a cada persona usuaria cuyo contenido se elimine o se suspenda la cuenta sobre el motivo de la eliminación o suspensión.

3. Apelación. Las empresas deben brindar una oportunidad significativa para apelar oportunamente cualquier eliminación de contenido o suspensión de la cuenta.

i. Transparencia

Sobre el principio de transparencia, los PSC establecen una serie de mínimos requeridos que las empresas deben dar a conocer para asegurar a la sociedad el respeto del derecho de acceso a la información y las garantías suficientes para supervisar que su moderación no afecte el derecho a la libertad de expresión de forma arbitraria o desproporcionada.

Concretamente, los PSC establecen que los intermediarios deben hacer públicas múltiples categorías de información estadística, incluyendo:

- Número total de publicaciones y cuentas marcadas (*flagged*).
- Número total de publicaciones eliminadas y cuentas suspendidas.
- Número de publicaciones y cuentas marcadas, y número de publicaciones eliminadas y cuentas suspendidas, por categoría de la regla que fue violada.
- Número de publicaciones y cuentas marcadas, y cantidad de publicaciones eliminadas y cuentas suspendidas, por formato de contenido en cuestión (por ejemplo, texto, audio, imagen, video, transmisión en vivo).
- Número de publicaciones y cuentas marcadas, y cantidad de publicaciones eliminadas y cuentas suspendidas, dependiendo de la fuente; es decir, Gobiernos, *revisores de confianza*, personas usuarias, diferentes tipos de detección automatizada.
- Número de publicaciones y cuentas marcadas, y cantidad de publicacio-



nes eliminadas y cuentas suspendidas, por ubicación de los revisores de confianza y personas usuarias afectadas (cuando sea evidente).

Los PSC recomiendan que los datos sugeridos deben proporcionarse en un informe regular, idealmente trimestralmente, en un formato legible para ser procesado por bases de datos y con licencia abierta.

ii. Notificación

La importancia de que las empresas notifiquen a las personas usuarias cuando sus expresiones deben de ser limitadas tiene como base una orientación detallada por parte de la empresa a la comunidad para que sus usuarias tengan conocimiento de qué contenido está prohibido. Se deben incluir ejemplos de contenido permitido e inadmisible y las pautas utilizadas por los revisores o moderadores. Las empresas también deben proporcionar una explicación de cómo se utiliza la detección automatizada en cada categoría de contenido.

Al proporcionar a una persona usuaria un aviso sobre por qué se ha eliminado su publicación o se ha suspendido una cuenta, un nivel mínimo de detalle para un aviso adecuado incluye:

- URL, extracto de contenido y otra información suficiente para permitir la identificación del contenido eliminado.
- La cláusula específica de las pautas que el contenido infringe.
- Cómo se detectó y eliminó el contenido (denunciado por otras personas usuarias, gobiernos, revisores de confianza, detección automatizada o quejas externas legales o de otro tipo). Por lo general, no se debe revelar la identidad de las usuarias individuales; sin embargo, el contenido señalado por el gobierno debe identificarse como tal.
- Explicación del proceso mediante el cual la persona usuaria puede apelar la decisión.

Los avisos deben estar disponibles en un formato duradero que sea accesible incluso si la cuenta de una persona usuaria se suspende o cancela. A las usuarias que denuncian contenido también se les debe presentar un registro del contenido que han informado y los resultados de los procesos de moderación.

iii. Apelación

Los PSC prevén que un mínimo -deseable- en las apelaciones consiste en que:

- El proceso sea claro y fácilmente accesible para las personas usuarias, incluyendo información sobre el tiempo de resolución y las maneras de dar seguimiento al proceso.
- Las apelaciones sean revisadas por una o por un panel de personas humanas que no estuviesen involucradas en la decisión inicial.
- La oportunidad de presentar información adicional que sea considerada en la revisión.
- La notificación de los resultados de la revisión y un pronunciamiento al respecto del razonamiento para tomar la decisión que permita a la persona usuaria entenderla.

Los PSC recomiendan que, a largo plazo, los procesos de revisión externa independientes también pueden ser un componente importante para que los usuarios puedan buscar una reparación.



E. RECOMENDACIONES EN TORNO A LA REGULACIÓN DE LA MODERACIÓN DE CONTENIDOS LLEVADA A CABO POR LOS INTERMEDIARIOS DOMINANTES EN INTERNET

A la luz de las estándares interamericanos sobre la libertad de expresión y en atención a la complejidad y obstáculos prácticos de la moderación de contenidos en línea que han sido desarrollados, consideramos indispensable que al contemplar generar esquemas de autorregulación, corregulación o regulación estatal en torno a la moderación de contenidos se tomen en cuenta las siguientes recomendaciones:

1. No responsabilidad de intermediarios por expresiones de terceras personas

Cualquier esquema regulatorio debe partir de la premisa general de que los intermediarios no deben ser responsabilizados por las expresiones de terceros en circunstancias en las que no han estado involucrados en la modificación de dicho contenido, pues de lo contrario se producen fuertes incentivos para una moderación de contenidos propensa a la censura de expresiones legítimas.

Lo anterior implica además que no deben imponerse obligaciones de monitoreo o filtrado proactivo de contenidos.

2. Enfoque diferenciado y delimitación clara de los intermediarios a los que les resultaría aplicable la regulación

Deben delimitarse con precisión los intermediarios a los cuales les resultaría aplicable cualquier regulación, asegurándose de que los parámetros

utilizados para definir a los sujetos obligados sean suficientemente estrictos para únicamente resultar aplicables a aquellos intermediarios que por su tamaño, número de usuarios, nivel de ingresos, cuota de mercado, o cualquier otro criterio que resulte relevante, posean la capacidad real para afectar significativamente el flujo informativo.

La adopción de un enfoque diferenciado y de una delimitación clara de los sujetos obligados es fundamental para evitar que la regulación tenga efectos anticompetitivos, es decir, que favorezca a los actores dominantes con mayor capacidad económica, técnica y administrativa para cumplir con la regulación, excluyendo o generando cargas desproporcionadas sobre intermediarios de menor tamaño o experiencia, provocando mayor concentración y afectando la pluralidad y diversidad en la oferta de servicios en Internet.

3. Políticas consistentes con los derechos humanos

Cualquier esquema de autorregulación, corregulación o de regulación estatal debe promover la adopción de políticas de contenido dirigidas a las personas usuarias que se encuentren apegados a los estándares interamericanos sobre derechos humanos.

En este sentido, la regulación estatal debe abstenerse de imponer obligaciones de remoción de contenidos, salvo en lo relativo a las categorías de discurso no protegido estrictamente definidas por el sistema interamericano. A saber, los contenidos que constituyen abuso sexual infantil, la incitación pública y directa al genocidio y la propaganda en favor de la guerra y la apología al odio nacional, racial o religioso que constituya incitación a la violencia.

Resulta particularmente perniciosa la imposición de obligaciones de moderación de contenidos a partir de categorías vagas e imprecisas, las cuales pueden generar un efecto inhibidor en las personas usuarias o inclusive conceder una amplia discrecionalidad para que el Estado y los actores privados restrinjan expresiones indebidamente.

Las políticas establecidas por los intermediarios respecto al contenido deben ser claras, precisas y accesibles para las personas usuarias. De esta manera las personas usuarias deben estar en posibilidad de comprender qué

tipos de contenido se encuentran prohibidos y serán removidos o sufrirán otro tipo de consecuencia, como la disminución de visibilidad (down ranking), la suspensión o la cancelación de la cuenta.

Los intermediarios deben aplicar de manera consistente y no discriminatoria sus políticas de contenido. Para ello, deben asegurarse de evaluar los métodos automatizados y no automatizados que sean utilizados en la moderación de contenidos de manera que sea posible detectar sesgos, errores o una calidad deficiente en la toma de decisiones.

Las compañías solamente deben utilizar procesos automatizados para identificar y remover contenidos o suspender cuentas cuando se utilicen en conjunción con mecanismos de revisión humana o exista una confianza suficientemente alta de la calidad y precisión de esos procesos.

En el diseño, implementación y evaluación de las políticas de contenido debe tenerse especial consideración respecto del impacto diferenciado que las políticas pueden producir sobre grupos específicos, en especial en razón de género, raza, idioma, discapacidad, edad, entre otros, así como en razón del contexto, como el electoral, de protesta social o de violencia ejercida por el Estado o por grupos de la delincuencia organizada.

4. Limitación de restricciones de contenido exigidas por el derecho de otros países no compatibles con las normas de derechos humanos interamericanas

Los Estados deben abstenerse de exigir que decisiones de remoción de contenidos sean aplicadas globalmente, en particular aquéllas que no sean compatibles con los estándares interamericanos de derechos humanos.

Los intermediarios a quienes les sean impuestas obligaciones legales de remoción de contenidos que no sean compatibles con los estándares interamericanos de derechos humanos deben esforzarse por combatir judicialmente dichos requerimientos y/o limitar geográficamente los efectos de dichas remociones de manera que las mismas no sean aplicadas a las personas usuarias ubicadas en países que forman parte del sistema interamericano.



5. Transparencia

Debe garantizarse la mayor transparencia posible sobre la toma de decisiones de moderación de contenidos por parte de los intermediarios. Los intermediarios deben publicar periódicamente información estadística suficiente para que las personas usuarias, investigadoras y la sociedad en general puedan evaluar los efectos de las decisiones de moderación de contenidos.

La información estadística publicada debe encontrarse desagregada en múltiples niveles y encontrarse estructurada en formatos abiertos. Los Principios de Santa Clara 2.0 ofrecen una guía detallada de la información estadística que debe publicarse en torno a la moderación de contenidos a partir de decisiones unilaterales por parte del intermediario, así como aquéllas en respuesta a un requerimiento de autoridad.

6. Notificación

Los esquemas de autorregulación, corregulación o regulación estatal deben asegurar que los intermediarios notifiquen a las personas usuarias afectadas directamente por una decisión de moderación de contenidos sobre la razones de dicha decisión.

Las notificaciones deben contener información suficiente para que la presunta infractora pueda evaluar la pertinencia o licitud de la decisión, incluyendo información que identifique con claridad el contenido supuestamente infractor, el fundamento normativo en el que se basa la decisión, el método de detección del contenido (si es automatizado, reportado por otros usuarios o solicitado por una autoridad) y en el caso de remociones motivadas por requerimientos de autoridad, el fundamento legal y la identificación de la autoridad requirente.

Las notificaciones deben ser oportunas, accesibles y establecer con claridad los mecanismos de apelación que se encuentran disponibles para la presunta infractora.

7. Apelación

Los esquemas de autorregulación, corregulación o regulación estatal deben disponer que los intermediarios establezcan mecanismos internos de apelación de las decisiones de moderación de contenidos.

Los mecanismos de apelación deben ser capaces de revertir una decisión de remoción de contenidos, de suspensión de una cuenta o cualquier otra acción derivada de la implementación de las políticas de contenidos del intermediario. En su caso, deben incorporar reparaciones de acuerdo al marco interamericano de derechos humanos para los casos en que afecten derechos humanos de personas usuarias en sus plataformas.

Los intermediarios deben aportar información accesible, oportuna y suficiente para que las personas usuarias afectadas por una decisión de moderación de contenidos o con un interés legítimo respecto de la misma puedan acceder a los mecanismos de apelación. Lo anterior implica conocer detalles sobre el proceso, los canales de comunicación para el seguimiento del mismo y el tiempo aproximado de su resolución, el cual debe ser el más breve posible, especialmente en contextos como el electoral o de protesta, en donde la demora puede hacer ineficaz la decisión dentro del proceso de apelación.

8. Desagregación de la toma de decisiones sobre la moderación de contenidos

Es recomendable que los intermediarios contemplen el establecimiento de mecanismos independientes para la revisión de las decisiones y políticas de moderación de contenidos. Por ejemplo, mecanismos como el Oversight Board de Facebook o los Consejos de Redes Sociales propuestos por la organización ARTICLE 19 pueden ayudar a evitar conflictos de interés y otorgar mayor legitimidad a las decisiones sobre moderación de contenidos de los intermediarios dominantes.

En la implementación de este tipo de mecanismos debe garantizarse la diversidad y participación equitativa de diversos grupos de la sociedad, incluyendo diversidad geográfica, de género, racial o étnica, entre otras categorías.

Los intermediarios deben garantizar la sostenibilidad financiera y la independencia de los mecanismos desagregados de toma de decisiones respecto de la moderación de contenidos, para lo cual dichos mecanismos deben actuar a la luz de los parámetros de transparencia previamente desarrollados.



9. Otras medidas para reducir la concentración y fomentar la pluralidad y diversidad en Internet

La concentración de poder de algunos intermediarios incrementa la trascendencia de sus decisiones de moderación de contenidos para el flujo informativo, por lo que resulta fundamental que los Estados implementen medidas para favorecer la competencia, la pluralidad y la diversidad en Internet.

La regulación y la labor de las autoridades de competencia deben conducir a la adopción de las medidas necesarias para evitar o sancionar conductas anticompetitivas por parte de los intermediarios dominantes en Internet, incluyendo medidas como la interoperabilidad y la desincorporación de activos.

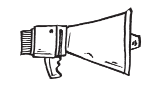
Igualmente, resulta fundamental garantizar el principio de neutralidad de la red y la prohibición de ofertas de tasa cero (*zero rating*) basadas en criterios comerciales que conlleven efectos anticompetitivos y de concentración en algunos intermediarios.

La adopción de medidas para reducir la concentración y fomentar la pluralidad y diversidad en Internet puede requerir la reformas legales o regulatorias o la realización de acciones de implementación efectiva de la regulación existente por parte de diversas autoridades reguladoras.

10. Participación multisectorial en la definición de políticas y evaluación de prácticas.

En la definición de los esquemas de autorregulación, corregulación y regulación estatal relacionados a la moderación de contenidos debe garantizarse la participación abierta de todas las partes interesadas, incluyendo a la sociedad civil, a la industria y a los organismos internacionales, de manera que puedan advertirse oportunamente deficiencias o insuficiencias en dichos esquemas.

A su vez, deben contemplarse mecanismos multisectoriales de seguimiento y evaluación de la moderación de contenidos implementada por los intermediarios.



LA MODERACIÓN DE CONTENIDOS DESDE UNA PERSPECTIVA INTERAMERICANA

Por: Vladimir Alexei Chorny Elizalde, Luis Fernando García Muñoz y Grecia Elizabeth Macias Llanas

Ciudad de México. México, Marzo 2022


AlSur
ALSUR.LAT
R3D
Red en Defensa
de los Derechos Digitales
R3D.MX